UNIVERSITY OF DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

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

# Runtime Analyses for a Simple Multi-objective Evolutionary Algorithm

Oliver Giel

No. CI-155/03

Technical Report ISSN 1433-3325 September 2003

Secretary of the SFB 531 · University of Dortmund · Dept. of Computer Science/XI
44221 Dortmund · Germany

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

# Runtime Analyses for a Simple Multi-objective Evolutionary Algorithm

Oliver Giel*

Fachbereich Informatik, LS 2, Univ. Dortmund
44221 Dortmund, Germany
oliver.giel@uni-dortmund.de

**Abstract.** Evolutionary algorithms are not only applied to optimization problems where a single objective is to be optimized but also to problems where several and often conflicting objectives are to be optimized simultaneously. Practical knowledge on the design and application of multi-objective evolutionary algorithms (MOEAs) is available but well-founded theoretical analyses can hardly be found. Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) have started such an analysis for two simple multi-objective evolutionary algorithms (SEMO and FEMO). In this work, the (expected) runtime of a variant of SEMO that searches globally is investigated. It is proven that the expected runtime is $O(n^n)$ for all objective functions $\{0,1\}^n \to \mathbb{R}^m$. For each $d \in \{2,\ldots,n\}$, a bicriteria problem such that the expected runtime is $\Theta(n^d)$ is presented. Bounds on the runtime and the expected runtime of the algorithm when applied to the problems LOTZ (leading ones trailing zeroes) and MOCO (multi-objective counting ones) are derived.

## 1 Introduction

Randomized search heuristics are applied to optimization problems in situations where problem-specific algorithms are not available. The lack of such algorithms can have various reasons. Problem-specific algorithms might be unknown for the considered problem, there might be not enough time and not enough experts to devise a problem-specific algorithm, or there might be only little knowledge about the structure of the problem. General search heuristics that do not employ problem-specific knowledge are of particular interest in theoretical investigations. In applications, these heuristics are often combined with problem-specific modules. Evolutionary algorithms (EAs) are such randomized search heuristics. They are not only applied to single-objective optimization problems but also to multi-objective optimization problems. Practical knowledge on the design and application of multi-objective evolutionary algorithms (MOEAs) has increased considerably in recent years but theoretical works are rare. A common approach to learn how EAs work is to analyze basic EAs. In this work, we analyze the expected runtime of a very simple but fundamental MOEA.

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

Theoretical analyses of the runtime of basic EAs in the scenario of single-objective optimization have been carried out in recent years. Most results giving time bounds consider discrete search spaces (e. g., Droste, Jansen, and Wegener (1998), Garnier, Kallel, and Schoenauer (1999), Droste, Jansen, and Wegener (2002), Wegener and Witt (2003)). Rigorous proofs on the runtime in a continuous search space have only been obtained recently (Jägersküpper (2003)). For an overview, refer to Wegener (2001) and Beyer, Schwefel, and Wegener (2002).

Works on the analysis of MOEAs have mostly focused on the limit behavior (convergence), i. e., the question under what conditions an algorithm can find the set of optimal solutions when time goes to infinity (Rudolph (1998a,b, 2001), Rudolph and Agapie (2000)). It is not possible to derive sharp bounds on the (expected) runtime without taking into account some properties of the function (the problem) to be optimized. Scharnow, Tinnefeld, and Wegener (2002) have analyzed the expected runtime of a variant of the (1+1) EA on a multi-objective formulation of the single-source shortest-path problem. However, the objectives of the problem are non-conflicting. Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) have been the first to analyze the (expected) runtime of two local search algorithms (SEMO and FEMO) for a problem with two conflicting objectives (LOTZ). Mutation is the only variation operator in these algorithms. In this work, we study the algorithm SEMO, but consider two different mutation operators. The first mutation operator flips a randomly chosen bit, the second mutation operator flips each bit independently. This results in two variants of SEMO which we call *local* SEMO and *global* SEMO, respectively. The local SEMO is the algorithm that has originally been named SEMO in Laumanns, Thiele, Zitzler, Welzl, and Deb (2002). The main concern of this work is the analysis the global SEMO even though the local SEMO will be considered as well.

The next section describes the scenario of multi-objective optimization in the framework of a partially ordered objective space and defines the goal of algorithms working in this scenario. Section 3 introduces the algorithms studied in subsequent sections and derives a tight bound on the expected runtime of the global SEMO in the worst case. Section 4 revisits the problem LOTZ (leading ones trailing zeros) from Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) but studies the global SEMO. Section 5 intensively studies the bicriteria problem MOCO (multi-objective counting ones) from Thierens (2003). It is shown that the runtimes of both algorithms are $\Theta(n^2 \log n)$ with high probability. Yet, the expected runtimes are much larger. Finally, in Section 6, the well-known test function $x \mapsto (x^2, (x-2)^2)$ is adapted to the Boolean decision space and bounds on the runtime are derived.

## 2 Scenario and Basic Definitions

In the scenario of multi-objective optimization, $m$ incommensurable and often conflicting objectives of a solution to some problem have to be optimized at the same time. The objective space $F$ can be thought of as a set of real-valued vectors such that each of $m$ components of a vector represents an objective of a solution. We assume all objectives to be maximized. Obviously, an objective vector $x$ is not better than another vector $y$ if each component of $x$ is not larger than the corresponding component of $y$. However, one cannot tell which of two distinct vectors is better in general. There is no natural total order on the objective space if the objectives are incommensurable. In this scenario, the aim of optimization is to find solutions such that an improvement regarding one objective can only be achieved at the expense of another objective. We follow Rudolph (1998a, 2001) for basic definitions.

**Definition 1 (preorder, partial order).** *Let $F$ be a set and $\preceq$ a binary relation in $F$. The relation $\preceq$ is called a* preorder *if it is reflexive and transitive. The pair $(F, \preceq)$ is called a* partially ordered set (poset) *if $\preceq$ is an antisymmetric ($\forall\, x, y \in F\colon x \preceq y \wedge y \preceq x \Rightarrow x = y$) preorder. Distinct $x$ and $y$ are* incomparable, *denoted $x \parallel y$, if neither $x \preceq y$ nor $x \succeq y$. Otherwise, if $x \preceq y$ or $x \succeq y$, $x$ and $y$ are* comparable. *In particular, $x$ is comparable to $x$.*

The relation in the set of real-valued vectors described above is a partial order.

**Definition 2 (domination and maximal elements).** *If $x \preceq y$, we say $y$* weakly dominates *$x$. We say $y$* dominates *$x$, denoted $x \prec y$, if $x \preceq y$ and $x \neq y$. An element $x^* \in F$ is called* maximal element *of the preordered set $(F, \preceq)$ if there is no $x \in F$ such that $x^* \prec x$. $M(F, \preceq)$ is the set of all maximal elements in $(F, \preceq)$.*

If $F$ is a finite set, the set $M(F, \preceq)$ is finite and complete. $M(F, \preceq)$ is said to be *complete* if for each $x \in F$ there exists an $x^* \in M(F, \preceq)$ such that $x \preceq x^*$.

In the framework of multi-objective optimization without constraints, we have the *decision space* $X$ (the set of all possible solutions), the partially ordered *objective space* $F$ (the poset of objective vectors), and an *objective function* $f\colon X \to F$. The aim of multi-objective optimization is *not* to compute the set of maximal elements in the objective space. We are rather interested in a set of best solutions in the decision space, the preimage of the maximal elements in the objective space. As the objective function is generally not a bijection, the preimage might be empty or considerably large. We must take care with regard to the definition of the aim in solving a multi-objective optimization problem.

**Definition 3 ($\preceq_f$).** *Let $X$ be the decision space and let $(F, \preceq)$ be the partially ordered objective space. Let $f\colon X \to F$ be a mapping. Then $f$ induces a preorder $\preceq_f$ on $X$ by the following definition:*

$$
\begin{aligned}
x \prec_f y \;&:\Leftrightarrow\; f(x) \prec f(y),\\
x =_f y \;&:\Leftrightarrow\; f(x) = f(y),\\
x \preceq_f y \;&:\Leftrightarrow\; x \prec_f y \,\vee\, x =_f y.
\end{aligned}
$$

In general, the preorder $\preceq_f$ is not a partial order since $x \preceq_f y \wedge y \preceq_f x \not\Rightarrow x = y$.

We use the notion of Pareto optimality if $f = (f_1, \ldots, f_m)$ is a vector-valued objective function, i.e., if $F$ is a subset of $\mathbb{R}^m$.

**Definition 4 (Pareto front, Pareto set).** *Let $X$ be a finite decision space, let $F := f(X) = \{f(x) \mid x \in X\} \subseteq \mathbb{R}^m$ be the objective space, and let the partial order $\preceq$ in $F$ be defined by*

$$(y_1, \ldots, y_m) \preceq (z_1, \ldots, z_m) \;\Leftrightarrow\; \forall i\colon y_i \leq z_i. \tag{1}$$

*The set of all maximal elements $F^* = M(F, \preceq)$ in the objective space is called* Pareto front*. An element $x \in X$ in the decision space is* Pareto optimal *if $f(x)$ belongs to the Pareto front $F^*$. The set of all Pareto optimal elements $X^* = f^{-1}(F^*)$ is called* Pareto set.

Definition 4 provides a surjective mapping $f$ and ensures that the objective space $(F, \preceq)$ is a finite poset with a finite (and complete) set of maximal elements, the Pareto front. In the following, we assume the scenario of Definition 4.

Roughly speaking, the goal of multi-objective optimization is to compute the Pareto set $X^*$. This goal can be too ambitious if the Pareto set is fairly large. However, if the Pareto set is large (e.g., exponential size) and the Pareto front is small (e.g., polynomial size) there are solutions $x^1, \ldots, x^k$ with the same objective value $f(x^1) = \cdots = f(x^k)$. In this case, a set of solutions should imply only one solution $x \in \{x^1, \ldots, x^k\}$. Provided that the Pareto front is not too large, a set $A \subseteq X^*$ representing each objective value in the Pareto front $F^*$ at least once is a reasonable set of solutions.

**Definition 5 (approximation set).** *A set $A' \subseteq F$ is called an* approximation set *(for the Pareto front) if no element in $A'$ is weakly dominated by any other element in $A'$ with respect to $\preceq$, i.e., any two distinct elements are incomparable.*

In Definition 5, we can replace weak domination by domination: For distinct elements in a poset, weak domination is equivalent to domination.

**Definition 6 (set of representatives).** *A* set of representatives *for a set $A' \subseteq F$ is a set $A \subseteq f^{-1}(A')$ such that $f(A) = A'$ and $|A| = |A'|$.*

In this work, the goal of an algorithm is to compute a set of representatives for the Pareto front. Clearly, if $f$ is not injective on the Pareto set, the Pareto set is not a set of representatives for the Pareto front. The computed set of solutions will be a subset of the Pareto set.

## 3 The Algorithm

The following evolutionary algorithm requires that the decision space is $X = \{0, 1\}^n$ and that there is a partial order relation $\preceq$ defined in the objective space $F = f(X)$. In particular, it applies to the scenario of multi-objective optimization in the Boolean decision space. The idea of the algorithm is that for each point of time $t$, the population $A_t$ is a set of representatives for the approximation set $f(A_t)$. The approximation set $f(A_t)$ is meant to approach the Pareto front $F^*$ as $t$ increases.

**Algorithm 1 (SEMO).**

> choose $x \in \{0,1\}^n$ uniformly at random
> determine $f(x)$
> $A \leftarrow \{x\}$
> **loop**
>   select $x \in A$ uniformly at random
>   create $x'$ from $x$ by mutation
>   determine $f(x')$
>   **if** $\forall\, z \in A\colon x' \not\preceq_f z$
>     $A \leftarrow \{z \in A \mid z \not\preceq_f x'\} \cup \{x'\}$
>   **end if**
> **end loop**

An implementation of the set $A$ needs to store search points $x$ together with their objective values $f(x)$. For the ease of notation, this aspect is not explicitly expressed in the description of Algorithm 1. Obviously, the initial population $A_1 = \{x_1\}$ is a set of representatives for the approximation set $\{f(x_1)\}$. The loop can be interpreted in the following way. At time $t$, the algorithm adds the offspring $x'_t$ to the population $A_t$ if there is no element in $A_t$ that weakly dominates $x'_t$, i.e., each element in $A_t$ is either dominated by $x'_t$ or incomparable to $x'_t$. If $x'_t$ is added to $A_t$, all elements in $A_t$ dominated by $x'_t$ are removed from $A_t$ at the same time, i.e., afterwards all elements of $A_t$ are incomparable with respect to $\preceq_f$. Hence, at each point of time $t$, $A_t$ is a set of representatives for the set $f(A_t)$. The latter set is an approximation set since for $x, y \in A_t$, $x \parallel_f y \Rightarrow f(x) \parallel f(y)$.

In applications, Algorithm 1 needs a stopping criterion. In this work, we are interested in the first point of time that the aim of the optimization process is reached and define the runtime of Algorithm 1 in the following way.

**Definition 7 (runtime).** *Let $A_t$, $t \in \mathbb{N}$, denote the population after the $(t-1)$th iteration of the loop, i.e., after $t$ objective function evaluations. The random number $T_f$ is the minimum $t$ such that $f(A_t) = F^*$. $T_f$ is called the* runtime *of Algorithm 1 for $f$.*

The assumption is that objective function evaluations are expensive and dominate the costs of all other operations in the loop. Then it is reasonable to call $T_f$ the runtime of Algorithm 1 for $f$. This measure is well accepted, particularly for evolutionary algorithms. It is also used in theoretical analyses of algorithms working in the black-box scenario (Droste, Jansen, Tinnefeld, and Wegener (2003)). From a practical point of view, the measure may not be fair if the population $A$ becomes very large. If $\Omega(|X|)$ elements of $(X, \preceq_f)$ are incomparable, the population may grow to size $\Omega(2^n)$. That means, the algorithm is only applicable if there is a much better bound for all $|A_t|$.

We have not specified the mutation operator yet. Depending on the mutation operator we refer to Algorithm 1 either as *local* SEMO or *global* SEMO. If the mutation operator flips only one bit chosen uniformly at random, we obtain the local search algorithm studied in Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) – there called SEMO (simple evolutionary multi-objective optimizer). In this paper, we call this algorithm *local* SEMO. If the mutation operator flips each bit independently with probability $1/n$ we call the algorithm *global* SEMO.

The local SEMO searches locally in the manner of a hill climber. If there is a subset of non-Pareto optimal points $\widehat{X}$ in the search space such that all Hamming neighbors of these points outside $\widehat{X}$ are (weakly) dominated by the points in $\widehat{X}$, the local SEMO's population cannot escape from $\widehat{X}$ if the entire population is contained in $\widehat{X}$. The following example problem with two objectives shows that this can happen with an overwhelming probability. The first objective is the number of ones in a solution $x$ if this number is even, otherwise it is 0. The second objective is the number of zeroes if this number is strictly less than $(1/4)n$, otherwise it is 0. Clearly, Pareto optimal solutions have at least $(3/4)n$ ones. By Chernoff bounds (e. g., Motwani and Raghavan (1995)), the probability of choosing an initial string $x$ with $i < (2/3)n$ ones is $1 - e^{-\Omega(n)}$, i. e., exponentially close to 1. If this happens and $i$ is odd, the objective value of $x$ is $(0, 0)$. The Hamming neighbors have objective values $(i - 1, 0)$ or $(i + 1, 0)$ and dominate $x$. One of them is created first and replaces $x$ in the population. Now, either by the initial step or by the first mutation, the algorithm is in the situation that the number of ones in the only individual $x$ in the population is even and at most $(2/3)n$. All Hamming neighbors of $x$ have objective values $(0, 0)$ and are dominated by $x$. Hence, the population gets stuck with an overwhelming probability before it reaches any point in the Pareto set. Therefore, SEMO has no finite expected runtime in the general case of an arbitrary $f$. In our example, restarts (multiple runs) do not help much since the probability of choosing a bad initial point is exponentially close to 1. The same applies to a variant of the local SEMO in Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) called FEMO (Fair Evolutionary Multi-objective Optimizer). We conclude that local search strategies can only be applied if we have some intuition of the optimization problem that suggests that such strategies are not very likely to get trapped.

In this paper, we focus on the global SEMO. In contrast to local search strategies, the population of the global SEMO will not get stuck in local optima forever. Local search strategies are typically easier to analyze than global search strategies. Nevertheless, the analysis of local search strategies can give insight into the problem at hand and is often a good starting point for the analysis of global search strategies. Sometimes it is easier to understand the main ideas of a proof when discussing the local SEMO first. Our first step is to study the expected runtime of the global SEMO in the worst case, i. e., the expected runtime if the objective function is chosen by an adversary. It is easy to see that for $n = 1$, the runtime of the global SEMO is at most 2. In the remainder of this paper, we assume that the dimension $n$ of the Boolean decision space is at least 2.

**Theorem 1.** *For any $f\colon \{0,1\}^n \to \mathbb{R}^m$, the expected runtime $E(T_f)$ of the global SEMO is bounded above by $(1+o(1))n^n$. There are functions $f$ where $E(T_f) \geq n^n$.*

The proof of Theorem 1 employs the following lemma.

**Lemma 1.** *Given a set $A$ of at least $n^{2\lceil \log n \rceil}$ points in $\{0,1\}^n$ and a point $x \in \{0,1\}^n \setminus A$. For more than half of the points in $A$, the Hamming distance to $x$ is at most $n - \lceil \log n \rceil$.*

*Proof.* The number of points $y \in \{0,1\}^n$ with Hamming distance $H(y,x) = k$ is $\binom{n}{k}$. The number of points with a Hamming distance to $x$ of at least $n-\lceil \log n\rceil+1$ is

$$\sum_{n-\lceil \log n\rceil+1 \leq k \leq n} \binom{n}{k} \leq \lceil \log n \rceil \binom{n}{\lceil \log n \rceil} \leq n^{\lceil \log n \rceil}$$

Consequently, $A$ contains at least $n^{2\lceil \log n\rceil} - n^{\lceil \log n \rceil} > (1/2)n^{2\lceil \log n \rceil}$ points $y$ with $H(x,y) \leq n - \lceil \log n \rceil$. □

*Proof (Proof of Theorem 1).* Whenever the global SEMO produces an offspring $y$ such that $f(y)$ is in the Pareto front $F^*$ and $f(y) \notin f(A)$, the algorithm adds $y$ to $A$. As $y$ is Pareto optimal, $y$ will never be removed from $A$. Before $A$ is a set of representatives for the Pareto front, such an individual $y$ exists. At time $t$, let $Y_t \subseteq X^*$ be the set of Pareto optimal decision vectors whose corresponding objective values are not yet represented by any decision vector in $A_t$. Formally, we define the target set $Y_t$ by $Y_t := X^* \setminus f^{-1}(f(A_t))$. The algorithm would accept each $y_j \in Y_t$ in the next mutation step. During a run of the algorithm, $Y_t \supset Y_{t+1}$ holds only if a new Pareto optimal point is added to $A_t$ and otherwise $Y_t = Y_{t+1}$. We define $A_0$ to be the empty population in the initialization step (Step 0) and, therefore, $Y_0 = X^*$. Let $X^* = Y_0 \supseteq \cdots \supseteq Y_{T_f} = \emptyset$ be the random sequence of target sets produced by the algorithm. Note that, for $k = |F^*|$, there are exactly $k+1$ mutually distinct sets $X^* = Y_{i_k} \supset \cdots \supset Y_{i_0} = \emptyset$ in this sequence and $|Y_{i_j}| \geq j$. Let $E(T_{i_j})$ denote the expected number of steps spent for the set $Y_{i_j}$. Then the expected runtime is

$$E(T_f) = \sum_{k \geq j \geq 2} E(T_{i_j}) + E(T_{i_1}).$$

The probability that the initial step selects a Pareto optimal search point is $\frac{|Y_{i_k}|}{2^n} \geq \frac{n-1}{n^n}(|Y_{i_k}|-1)$. For $t \geq 1$, let $x \in A_t$ denote the individual selected for mutation and let $Y_t = \{y_1, \ldots, y_{|Y_t|}\}$ be the target set at time $t$. There is at most one $y_j \in Y_t$ such that the Hamming distance $H(x,y_j) = n$, namely if $y_j = \overline{x}$. In all other cases, $H(x,y_j)$ is at most $n-1$. Hence, the probability that the algorithm creates an offspring in $Y_t$ is lower bounded by

$$\begin{aligned} &\sum_{1 \leq j \leq |Y_t|} (1/n)^{H(x,y_j)}(1-1/n)^{n-H(x,y_j)} \\ \geq &\sum_{1 \leq j \leq |Y_t|-1} (1/n)^{n-1}(1-1/n) \ = \ \frac{n-1}{n^n}(|Y_t|-1), \end{aligned}$$

and for $|Y_{i_j}| \geq 2$ the expected value $E(T_{i_j})$ is at most $\frac{n^n}{(n-1)(|Y_{i_j}|-1)}$. Since $|Y_{i_j}| \geq j$, we have

$$E(T_f) \leq \frac{n^n}{n-1} \sum_{k \geq j \geq 2} \frac{1}{j-1} + E(T_{i_1}). \tag{2}$$

Now we estimate the right-hand side of the last equation according to two cases.

The first case is $k \geq n^{2\lceil \log n \rceil} + 1$. We consider the steps when the target set is $Y_{i_1}$, i.e., the algorithm has discovered $k-1$ representatives for $k-1 = |F^*|-1$ points of the Pareto front before. Hence, $|A| = k-1 \geq n^{2\lceil \log n \rceil}$, and there is only one point $x'$ in the Pareto front that is not in $f(A)$. Let $x \in f^{-1}(x') = Y_{i_1}$. By Lemma 1, the probability that the algorithm selects an individual in $A$ such that the Hamming distance to $x$ is at most $n - \lceil \log n \rceil$ is at least $1/2$. Thus, $E(T_{i_1})$ is bounded above by

$$\left((1/2)(1/n)^{n-\lceil \log n \rceil}(1-1/n)^{\lceil \log n \rceil}\right)^{-1} \leq 2en^{n-\lceil \log n \rceil}.$$

Using $k = |F^*| \leq |X^*| \leq 2^n$, we can upper bound (2) by

$$\frac{n^n}{n-1} \sum_{j=2}^{2^n} \frac{1}{j-1} + 2en^{n-\lceil \log n \rceil} \leq n^n \left( \frac{H_{2^n-1}}{n-1} + \frac{2e}{n^{\lceil \log n \rceil}} \right).$$

The last expression is strictly smaller than $n^n$ for $n$ large enough since the harmonic number $H_{2^n-1}$ is bounded by $\ln(2^n-1)+1 \leq 0.7n+1$.

The second case is $k \leq n^{2\lceil \log n \rceil}$. When the target set is $Y_{i_1}$, the probability that the next mutation step creates a point in this set is at least $1/n^n$. In the initial step, the probability is at least $k/2^n \geq 1/n^n$. Hence, $E(T_{i_1}) \leq n^n$, and (2) is bounded by

$$\frac{n^n}{n-1} \sum_{j=2}^{k} \frac{1}{j-1} + n^n = n^n \left( \frac{H_{k-1}}{n-1} + 1 \right) = (1+o(1))n^n,$$

using $H_{k-1} \leq \ln(k-1)+1 \leq 0.7 \log n^{2\lceil \log n \rceil} + 1 \leq 1.4\lceil \log n \rceil^2 + 1$.

For the lower bound, we consider the function

$$f(x) = \left( \prod_{1 \leq i \leq n} x_i, \prod_{1 \leq i \leq n} (1-x_i) \right).$$

Obviously, the objective space is $F = \{(0,0),(1,0),(0,1)\}$ and only $(0,0)$ is not maximal. Let $x$ be the Pareto optimal decision vector found first by the algorithm, i.e., either $x = 0^n$ or $x = 1^n$. Since $x$ dominates all decision vectors found before, the population now is $A = \{x\}$. The Hamming distance to the second Pareto optimum $\overline{x}$ is $n$. Hence, the expected waiting time is $n^n$. $\square$

Theorem 1 states a $\Theta(n^n)$ bound in the worst case. Note that the upper bound is independent of the number of objectives $m$ and that the lower bound is obtained from a bicriteria problem. In Section 6 we will present for each $d \in \{2, \dots, n\}$ a bicriteria problem such that the expected runtime of the global SEMO is $\Theta(n^d)$. The conclusion is that if we only consider expected runtimes, bicriteria problems are not easier for the local SEMO than problems with a high-dimensional objective space. There are bicriteria problems among the hardest problems. The scenario of multi-objective optimization includes the scenario of single-objective optimization. The (1+1) EA is perhaps the most fundamental evolutionary algorithm for single-objective optimization in the Boolean decision space $\{0,1\}^n$. Interestingly, it has the same expected runtime $\Theta(n^n)$ in the worst case (Droste, Jansen, and Wegener (2002)). If applied to a monocriteria problem, the global SEMO behaves almost like the (1+1) EA. One can also obtain $\Omega(n^n)$ bounds from some monocriteria problems that have been analyzed for the (1+1) EA, e.g., the problem DISTANCE considered in Droste, Jansen, and Wegener (2002).

## 4 LOTZ – Leading Ones Trailing Zeroes

The LOTZ function has been studied in Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) for the local SEMO and a variant of the local SEMO called FEMO. FEMO employs the same mutation operator as the local SEMO. Flipping exactly 1 bit in each step simplifies the analysis of these algorithms for this function. The effect is that the population size is bounded by 1 until the first point in the Pareto set is discovered, and then the algorithms explore the Pareto set without accepting solutions that are not Pareto optimal. Both properties do not carry over to the global SEMO. The selection mechanism of FEMO has been adapted to the LOTZ function, and in fact FEMO performs better on LOTZ. The expected runtimes for the local SEMO and FEMO are $\Theta(n^3)$ and $\Theta(n^2 \log n)$, respectively. In this section, we show that using independent bit flips with SEMO (i.e., the global SEMO) does not increase the runtime substantially. Moreover, the runtime is $O(n^3)$ with a probability exponentially close to 1. It is not known whether independent bit flips increase the runtime of FEMO for LOTZ.

**Definition 8.** *The functions* $\mathrm{LO}, \mathrm{TZ}\colon \{0,1\}^n \to \mathbb{N}$ *and* $\mathrm{LOTZ}\colon \{0,1\}^n \to \mathbb{N}^2$ *are defined by*

$$\begin{aligned} \mathrm{LO}(x) &:= \sum_{i=1}^{n} \prod_{j=1}^{i} x_j, \\ \mathrm{TZ}(x) &:= \sum_{i=1}^{n} \prod_{j=i}^{n} (1 - x_j), \\ \mathrm{LOTZ}(x) &:= \big(\mathrm{LO}(x), \mathrm{TZ}(x)\big). \end{aligned}$$

$\mathrm{LO}(x)$ is the number of leading ones in $x$ and $\mathrm{TZ}(x)$ the number of trailing zeroes. We define the relation $\preceq$ in $\mathbb{N}_0^2$ according to (1) and consider the partially ordered objective space $(\mathbb{N}_0^2 \cap \mathrm{LOTZ}(\{0,1\}^n), \preceq)$ and the preordered decision space $(\{0,1\}^n, \preceq_{\mathrm{LOTZ}})$. In the remainder of this section, we omit the subscript "LOTZ" in our notation.

**Proposition 1.** *The Pareto front $F^*$ is the set $\{(i, n-i) \mid 0 \le i \le n\}$, and the Pareto set $X^*$ is the set of all strings $1^i 0^{n-i}$, $0 \le i \le n$. The Pareto set $X^*$ is the only set of representatives for the Pareto front $F^*$.*

*Proof.* The set $\{(i,j) \mid 0 \le i+j \le n,\ i+j \ne n-1\}$ is the objective space, and only the elements $(i, n-i)$, $0 \le i \le n$, are not dominated by any other element. Obviously, $\mathrm{LOTZ}^{-1}(i, n-i)$ is the singleton set $\{1^i 0^{n-i}\}$. □

**Proposition 2.** *Let $A$ be a set of representatives for an approximation set $A'$. The cardinality of $A$ is at most $n+1$. If $A \ne X^*$, the cardinality of $A$ is at most $n$.*

*Proof.* As $|A| = |A'|$, it suffices to show $|A'| \le n+1$. The characteristic function of the objective space $F \subseteq \{0, \ldots, n\}^2$ can be viewed as a triangular matrix with 1-entries at $(i,j)$, $0 \le i+j \le n$ and $i+j \ne n-1$. The row index $i$ gives the number of leading ones, the column index $j$ the number of trailing zeroes. Since $A'$ is an approximation set, i.e., no element in $A'$ is dominated by any other element in $A'$, there is at most one element from each of the $n+1$ rows in $A'$. The same applies to columns. This shows that $|A'| \le n+1$. Assume $|A'| = n+1$. Then $A'$ chooses exactly one element in each row and each column. Hence, the characteristic function of $A'$ can be viewed as a permutation matrix. As $A' \subseteq F$, the 1-entries in the permutation matrix are also 1-entries in the triangular matrix representing $F$. It is easy to see that there is only one choice for $A'$, namely all elements placed on the diagonal $(i, n-i)$, $0 \le i \le n$. Hence, $A' = F^*$. □

**Theorem 2.** *The expected runtime of the global SEMO for* LOTZ *is $O(n^3)$. The runtime is $O(n^3)$ with a probability $1 - e^{-\Omega(n)}$.*

*Proof.* As the Pareto set $X^*$ is the unique set of representatives for $F^*$ (Proposition 1), the population becomes static if $A = X^*$. It can change at any time before this event happens. We discern two epochs in a typical run of the algorithm. The first epoch starts after the initialization and is finished by the step producing the first individual $x \in X^*$. The following epoch lasts until $A = X^*$.

First we show that a phase of $s := \lceil en^3 \rceil$ steps finishes the first epoch with a probability $1 - e^{-\Omega(n)}$. We consider the initial individual $x^0$ and the (random) sequence of individuals $x^1, x^2, x^3, \ldots$ in the first epoch such that $x^{i+1}$ causes $x^i$ to leave the population (because $x^{i+1} \succ x^i$). When the dominating individual $x^{i+1}$ is created, either the number of leading ones compared to $x^i$ is increased and the number of trailing zeros compared to $x_i$ is not decreased or vice versa. That implies that there are at most $n$ individuals in the above sequence that starts with $x^1$. If an offspring dominates its parent then it will be accepted and replace the parent individual. We estimate the probability to create $x^{i+1}$ in the next step by the probability that $x^i$ is chosen for mutation and the algorithm flips either only the leftmost 0 or only the rightmost 1. We call this event a success. Using

Proposition 2, the probability of a success is at least $(1/n)\cdot 2\cdot(1/n)\cdot(1-1/n)^{n-1} \geq 2/(en^2)$. Within the first phase, the expected number of successes is at least $2n$. By Chernoff bounds, the probability of less than $n$ successes is $e^{-\Omega(n)}$. The first phase of $s$ steps finishes the first epoch with a probability exponentially close to 1. To obtain an upper bound on the expected number of steps, we observe that our estimations also hold if we start a new phase with a population of up to $n$ non-optimal solutions. The expected number of phases is upper bounded by 2. This implies that the expected number of steps in the first epoch is $O(s) = O(n^3)$.

Next we show that, starting with at least one Pareto optimal element in $A$, after a phase of $s' = \lceil 2en^3 \rceil$ steps, $X^* = A$ with a probability $1 - e^{-\Omega(n)}$. The Pareto set can be viewed as a path from $0^n$ to $1^n$ that visits all strings $1^i 0^{n-i}$, $0 \leq i \leq n$. Obviously, each individual on the path has at least one Hamming neighbor on the path. As long as $A \neq X^*$, there exists at least one $x \in A$ with a Hamming neighbor $x' \in X^* \setminus A$ and by Proposition 2, $|A| \leq n$ holds. The probability of creating $x'$ in the next step is at least $(1/n)\cdot(1/n)\cdot(1-1/n)^{n-1} \geq 1/(en^2)$. Within $s'$ steps of a phase, the expected number of such successes is at least $2n$. Using Chernoff bounds again, the probability of less than $n$ successes is $e^{-\Omega(n)}$. Analogously to the first epoch, the expected number of steps is $O(n^3)$, too.

Combining the results for both epochs yields the bounds in the theorem. □

As mentioned before, an expected runtime of $\Theta(n^3)$ has been proved by Laumanns, Thiele, Zitzler, Welzl, and Deb (2002) for the local SEMO. The authors have also proved that the runtime is $\Omega(n^3)$ with probability $1 - e^{-\Omega(n)}$. In the proof of Theorem 2, we have only considered mutation steps where solely one bit flips. Therefore, our lower bounds on the probability of a success in the first and second epoch also hold for the local SEMO. We obtain that the runtime of the local SEMO is $O(n^3)$ with a probability $1 - e^{-\Omega(n)}$. Combining both results yields the following corollary. Additionally, we obtain an alternative proof for the $O(n^3)$ bound of the expected runtime of the local SEMO.

**Corollary 1.** *The expected runtime of the local SEMO for* LOTZ *is* $\Theta(n^3)$. *The runtime is* $\Theta(n^3)$ *with a probability* $1 - e^{-\Omega(n)}$.

## 5 MOCO – Multi-objective Counting Ones

The OneMax function, also called CountingOnes function, counts the number of ones in a bitstring $x = x_1, \ldots, x_n$, and the aim is to maximize the number of ones. We denote this function by $\|x\|$. The following bicriteria problem MOCO (multi-objective counting ones) has been introduced in Thierens (2003).

**Definition 9.** *For* $\varphi(x) = 2\pi\frac{\|x\|}{n}$ *and* $n = 4k$, *the function* $\text{MOCO}\colon \{0,1\}^n \to [-1,1]^2$ *is defined by*

$$\text{MOCO}(x) = \big(\cos\varphi(x), \sin\varphi(x)\big).$$

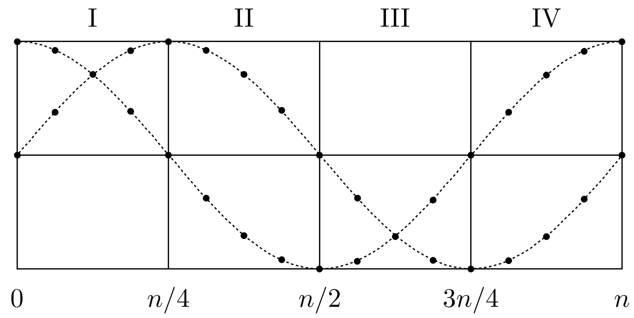


**Fig. 1.** MOCO for $n = 16$.

Based on approximate calculations and experimental results, Thierens (2003) conjectures that the expected runtimes of both the local and global SEMO are of the order of $n^2 \log n$. We prove that the expected runtime of the local SEMO is non-finite and the expected runtime of the global SEMO is $n^{\Omega(n)}$. Considering expected runtimes, MOCO is almost a worst-case example for both algorithms. Yet, the empirical study in Thierens (2003) suggests that the observed runtimes are typically polynomial and presumably of the order of $n^2 \log n$. The main goal of this section is to formalize this conjecture and to present corresponding proofs.

We prepare our analysis of the runtimes by considering a partitioning of the decision space into four regions (Fig. 1) as defined in Thierens (2003). The condition $n = 4k$ has been assumed for the ease of presentation: The objectives take the same values on rising and falling edges of the sine and cosine function, and, importantly, they hit the maxima and minima. Otherwise the following partitioning has to be defined more carefully. (The same or a similar condition is implicitly assumed in Thierens (2003).)

**Region I ($0 \leq \|x\| < n/4$):** Both sine and cosine are non-negative in this region. For each search point in Regions II–IV, at least one objective is non-positive. Therefore, no point in the region is dominated by any point in Regions II–IV. The second objective (sin) strictly increases and the first objective (cos) strictly decreases with the number of ones. Hence, for any two points $x$ and $y$ in this region, either $x =_{\text{MOCO}} y$ or $x \parallel_{\text{MOCO}} y$ holds. Therefore, all points in Region I belong to the Pareto set.

**Region II ($n/4 \leq \|x\| < n/2$):** In this region, a point with $i$ ones is dominated by any point with less than $i$ ones in Region II. Hence, any point with $n/4$ ones is a maximal element in this region. It is also a maximal element in the entire decision space $X$, i.e., a Pareto optimum.

**Region III ($n/2 \leq \|x\| < 3n/4$):** For all points $x$ and $y$ in this region, either $x =_{\text{MOCO}} y$ or $x \parallel_{\text{MOCO}} y$ holds. Every point in this region is dominated by every point in Region I. Hence, there are no Pareto optima here.

**Region IV ($3n/4 \leq \|x\| \leq n$):** In this region, a point with $i$ ones is dominated by every point with more than $i$ ones. Hence, the point with $n$ ones is the only maximal element in this region. Since the point with $n$ ones and the point with $n$ zeroes in Region I have the same objective values, the first point is a Pareto optimum.

The following proposition summarizes our discussion of the four regions.

**Proposition 3.** *The Pareto set $X^*$ is the set $\{x \mid 0 \leq \|x\| \leq n/4 \ \vee\ \|x\| = n\}$. The Pareto front $F^*$ is the set $\{(\cos\varphi(i), \sin\varphi(i)) \mid 0 \leq i \leq n/4\}$.*

The next proposition implies that the population size of the local and global SEMO is bounded by $n/4+1$ at any time. This bound is tight as the cardinality of the Pareto front is $n/4+1$. One can certainly give better bounds for specific situations, however, the use of better estimates would not improve our results substantially.

**Proposition 4.** *Let $A$ be a set of representatives for an approximation set $A'$. The cardinality of $A$ is at most $n/4+1$.*

*Proof.* We only have to show $|A'| \leq n/4+1$. Our partitioning of the decision space induces a partition of the objective space if we define that the objective vector $(1, 0)$ belongs to Region I. Each region contains $n/4$ objective vectors. For convenience, we use the same names for corresponding regions in the decision space and objective space.

Any two objective vectors in Region II are comparable. Hence, there is at most one vector from Region II in $A'$, and the same applies to Region IV. Now we distinguish two cases. The first case is that there is no vector from Region I in $A'$. There are at most $n/4$ vectors from Region III, and there is at most one additional vector from Region II and Region IV each. As any vector from Region II dominates at least the vector $(-1, 0)$ in Region III, there are at most $n/4+1$ objective vectors in $A'$. The second case is that there is at least one vector from Region I in $A'$. As each of the vectors in Region I dominates every vector in Region III, $A'$ contains no vector from Region III and at most $n/4$ from Region I. There is at most one additional vector from Region II and at most one additional vector from Region IV in $A'$. If there is a vector from Region IV, it excludes at least the vector $(1, 0)$ in Region I from $A'$. □

**Theorem 3.** *The local SEMO has no finite expected runtime for* MOCO*, and the expected runtime of the global SEMO is $n^{\Omega(n)}$.*

*Proof.* For the first statement, let the random variable $T$ denote the runtime of the local SEMO. Then $E(T)$ equals

$$\sum_{t\geq 0} t \cdot \mathrm{Prob}(T=t) \ = \ \sum_{t\geq 1} \mathrm{Prob}(T \geq t). \tag{3}$$

There is a positive probability of $2^{-n}$ that the initial point has $n$ ones. Starting there, the mutation operator can only create offspring with $n-1$ ones, which are dominated by the initial point. Therefore, the population cannot cross the border to Region III. Thus, $\mathrm{Prob}(T \geq t)$ is lower bounded by $2^{-n}$ for all $t$. Hence, the right-hand side of (3) diverges. We remark that any initial point with more than $3n/4$ ones leads to the same result.

For the second statement, let $T$ now denote the runtime of the global SEMO, and let $A_i$ denote the event that the initial search point has exactly $i$ ones. By Theorem 1, $E(T)$ is finite. It can be bounded below in the following way.

$$E(T) \;=\; \sum_{0\le i\le n} E(T \mid A_i)\cdot \mathrm{Prob}(A_i) \;\ge\; E(T \mid A_n)\cdot \mathrm{Prob}(A_n). \qquad (4)$$

Now we assume that the algorithm has selected an initial point with $n$ ones. The point is Pareto optimal and weakly dominates all points with at least $n/2$ ones. Hence, the mutation step has to flip more than $n/2$ ones into zeroes in order to generate a new point that will be accepted. The probability of this event in a single mutation step is

$$\sum_{\frac{n}{2}+1\le k\le n} \binom{n}{k}\left(\frac{1}{n}\right)^{k}\left(1-\frac{1}{n}\right)^{n-k} \le \sum_{0\le k\le n}\binom{n}{k}\left(\frac{1}{n}\right)^{n/2} \le 2^n n^{-n/2} = n^{-\Omega(n)},$$

implying $E(T \mid A_n) = n^{\Omega(n)}$. Hence, the right-hand side of (4) is $n^{\Omega(n)}\cdot 2^{-n} = n^{\Omega(n)}$. □

If $n$ is not very small, the expected runtimes of the local and global SEMO do not reflect the behavior observed in practice. Therefore, we are seeking for results that bound the probability that a run finds a solution in a certain time. We shall see that the runtimes of both variants of SEMO are $\Theta(n^2 \log n)$ with high probability.

Once there is a search point $x$ with $\|x\| = i$ in the population $A$, no other search point with $i$ ones can enter $A$ because some element of $A$ weakly dominates it. This holds true even if $x$ is removed from $A$ later. Thus, throughout a run of the local or global SEMO, at most one individual with $i$ ones, $0 \le i \le n$, occurs in $A$. In the remainder of this section, we use the following notation. For a specific run, let $x^i$ denote this individual with $i$ ones (if such an individual occurs in $A$ in the run). At any time, $x^{\mathrm{low}}$ denotes the individual with the least number of ones in the current population, and $x^{\mathrm{high}}$ denotes the individual with the largest number of ones. Clearly, $x^{\mathrm{low}}$ and $x^{\mathrm{high}}$ change with time and depend on the random decisions of the algorithm.

**Theorem 4.** *For any polynomial $p(n) > 0$, the runtime of the local SEMO for* MOCO *is $O(n^2 \log n)$ with a probability of at least $1 - 1/p(n)$.*

*Proof.* In this proof, we partition a run of the algorithm into three consecutive epochs and show that each of the first two epochs is finished successfully in time $O(n^2)$ with a probability $1 - e^{-\Omega(n)}$. Let $d$ denote the degree of the polynomial $p(n)$. We prove that the final epoch takes $O(n^2 \log n)$ steps with a probability of at least $1 - 1/n^{d+2}$ for $n$ large enough. Taking into account the failure probabilities in the first two epochs, the overall success probability is at least $1 - 1/n^{d+1} > 1 - 1/p(n)$ for $n$ large enough.

The first epoch lasts until the first individual with less than $n/2$ ones is contained in the population. We show that the first $O(n^2)$ steps produce such a point with probability $1 - e^{-\Omega(n)}$; otherwise we consider the run unsuccessful. The expected number of ones of the initial search point equals $n/2$, and large deviations from the expected value are unlikely: By Chernoff bounds, the initial point has less than $(9/16)n$ ones with a probability $1 - e^{-\Omega(n)}$. There is nothing to show if the initial point has less than $n/2$ ones. So we only consider the case that the number of ones is at least $n/2$ and less than $(9/16)n$. We name a step of the algorithm *relevant* if $x^{\text{low}}$ or $x^{\text{high}}$ is selected for mutation. By Proposition 4, a step is relevant with a probability of at least $1/(n/4 + 1)$, and in the first $(1/8)n^2 + (1/2)n$ steps subsequent to the initial step, the expected number of relevant steps is at least $n/2$. By Chernoff bounds, there are at least $\lfloor (6/16)n \rfloor$ relevant steps with probability $1 - e^{-\Omega(n)}$. Now we consider the phase containing the first $\lfloor (6/16)n \rfloor$ relevant steps (including the non-relevant steps inbetween) that starts right after the initial step. The phase is finished prematurely if an offspring with less than $n/2$ ones is created. Non-relevant steps never decrease $\|x^{\text{low}}\|$ or increase $\|x^{\text{high}}\|$ because the algorithms flips only a single bit. Only relevant steps may do so. Clearly, in the phase, the condition $n/2 \leq \|x^{\text{low}}\| \leq \|x^{\text{high}}\| < (15/16)n$ holds. By our discussion of Region II, III, and IV, it is clear that no search point with at least $(9/16)n$ ones and at most $(15/16)n$ ones dominates the initial point, and that every relevant step of the phase that decreases $\|x^{\text{low}}\|$ produces an offspring that enters the population. Given that a step is relevant, the probability that it selects $x^{\text{low}}$ and flips a one into a zero is at least $1/4$ because the algorithm selects individuals uniformly and $x^{\text{low}}$ has at least $n/2$ ones. By Chernoff bounds, $\|x^{\text{low}}\|$ decreases by at least $(1/16)n$ to less than $n/2$ in $\lfloor (6/16)n \rfloor$ relevant steps with a probability $1 - e^{-\Omega(n)}$.

The second epoch lasts until the first Pareto optimal point with at most $n/4$ ones is created. For each search point with less than $n/2$ ones, the second objective (sin) is positive, and for each point with at least $n/2$ ones in Region III and IV, the second objective is non-positive. Therefore, no point with at least $n/2$ ones weakly dominates any point with less than $n/2$ ones. Thus, if a step in this epoch selects $x^{\text{low}}$ and flips a one, it produces an offspring that replaces $x^{\text{low}}$ and becomes $x^{\text{low}}$ in the new population. Since $\|x^{\text{low}}\| \geq (1/4)n$, the probability of this event is at least $1/(n/4+1) \cdot (1/4) = 1/(n+4)$. In the next $(1/2)n^2 + 2n$ steps, the expected number of such events is at least $n/2$. By Chernoff bounds, the probability of at least $n/4$ such events is $1 - e^{-\Omega(n)}$. Thus the duration of the second epoch is $O(n^2)$ with the desired probability.

Unless $x^i$ is the initial individual, $x^i$ is an offspring either of $x^{i+1}$ or $x^{i-1}$. Let $x^\ell$ be the individual with at most $n/4$ ones that appears in $A$ first. The aim of the third epoch is to produce the points $x^{\ell+1}, \ldots, x^{n/4}$ and the points $x^{\ell-1}, \ldots, x^0$. (We pessimistically assume that the search point $x^0$ has to be produced.) For $\ell + 1 \leq i \leq n/4$, the probability that the next step produces the next $x^i$ is lower bounded by $1/(n/4 + 1) \cdot (n - i)/n \geq 3/(n + 4)$. By Chernoff bounds, the next $O(n^2)$ steps produce all $x^{\ell+1}, \ldots, x^{n/4}$ with probability $1 - e^{-\Omega(n)}$. Let $I$ be the set of the $\ell$ bit positions of the ones in $x^\ell$. Now we call a step *relevant* if the algorithm selects the individual $x^{\text{low}} = x^i$, for some $1 \leq i \leq \ell$, for mutation. Whenever a relevant step mutates a new bit position in $I$, $x^{i-1}$ is created. Now the scenario is similar to the coupon collector's problem. The probability that the next step is relevant and mutates a specific position in $I$ is at least $1/(n/4+1) \cdot (1/n) = 4/(n^2+4n) =: p$. The probability that position $j \in I$ has not been flipped in a relevant step in a phase of $r := ((d+3)/p) \ln n = O(n^2 \log n)$ steps is at most $(1-p)^r \leq e^{-(d+3)\ln n} = n^{-(d+3)}$. The probability of the union of these events is at most $\sum_{i=1}^{n/4} n^{-(d+3)} \leq (1/4) \cdot n^{-(d+2)}$. Taking into account the failure probability $e^{-\Omega(n)}$ for $x^{\ell+1}, \ldots, x^{n/4}$, the failure probability of this epoch is bounded by $n^{-(d+2)}$, for $n$ large enough. □

Our upper bound $O(n^2 \log n)$ for the local SEMO holds with a high probability, i. e., polynomially close to 1. The next theorem states that the lower bound $\Omega(n^2 \log n)$ holds even with an overwhelming probability, i. e., exponentially close to 1.

**Theorem 5.** *For every constant* $\varepsilon < 1$*, the runtime of the local SEMO for* MOCO *is* $\Omega(n^2 \log n)$ *with a probability* $1 - e^{-\Omega(n^\varepsilon)}$.

*Proof.* By Chernoff bounds, the initial search point has more than $n/4$ ones with a probability of at least $1-e^{-\Omega(n)}$. Starting with an initial point with more than $n/4$ ones, there might be a chance that the image of the population does not converge to the Pareto front. For our lower bound, we can ignore this case and assume that at some point of time, the first individual with at most $n/4$ ones is created. As the algorithm flips only 1 bit in each step, it will be an individual with exactly $n/4$ ones. The remaining runtime of the algorithm is lower bounded by the time to produce the individuals $x^{n/4-1}, \ldots, x^1$. (Note that we must not include $x^0$ because there is a chance that $x^n$ is created.) A step of the algorithm can only produce the next individual in this series if it selects the individual with the lowest number of ones and flips a one into a zero. Let the time $t = 0$ be the step in which $x^{\lfloor n/8 \rfloor}$ is created. At that time, the population contains more than $n/8$ Pareto optima, namely $x^{\lfloor n/8 \rfloor}, \ldots, x^{n/4}$. We call a step at time $t > 0$ *relevant* if it selects $x^{\text{low}}$ for mutation. Let $I$ be the set of the $\lfloor n/8 \rfloor$ bit positions of the ones in $x^{\lfloor n/8 \rfloor}$. W. l. o. g. let $I = \{1, \ldots, \lfloor n/8 \rfloor\}$. The next Pareto optimum is only created if a relevant step mutates a new position in $I$, i. e., a position that has not been flipped in a relevant step prior to the current step. Now we consider a phase of $s := \lfloor (1-\varepsilon)(n^2/8-1) \ln n \rfloor = \Omega(n^2 \log n)$ steps that start at time $t = 1$. Our aim is to estimate the probability that the phase is successful, i. e., that it produces $x^{\lfloor n/8 \rfloor - 1}, \ldots, x^1$. Call this event $S$. The situation is similar to a coupon collector's problem where all but one coupon have to be collected. Let $A_i$, $1 \leq i \leq \lfloor n/8 \rfloor$, denote the event that the bit at position $i \in I$ has flipped at least once in a relevant step of the phase. Then $S$ is the event that at least $\lfloor n/8 \rfloor - 1$ of the events $A_1, \ldots, A_{\lfloor n/8 \rfloor}$ occur, i. e., $\text{Prob}(S)$ equals

$$\text{Prob}\left((A_2 \cap \cdots \cap A_{\lfloor n/8 \rfloor}) \cup \cdots \cup (A_1 \cap \cdots \cap A_{\lfloor n/8 \rfloor - 1})\right).$$

As the probability of the union of events is upper bounded by the sum of the probabilities of each event, and each event has the same probability, we have

$$\begin{aligned}\mathrm{Prob}(S) &\le \lfloor n/8\rfloor \,\mathrm{Prob}(A_1 \cap \cdots \cap A_{\lfloor n/8\rfloor -1})\\ &= \lfloor n/8\rfloor \,\mathrm{Prob}(A_1)\,\mathrm{Prob}(A_2 \mid A_1)\,\mathrm{Prob}(A_3 \mid A_1 \cap A_2)\cdots\\ &\quad\ \mathrm{Prob}(A_{\lfloor n/8\rfloor -1} \mid A_1 \cap \cdots \cap A_{\lfloor n/8\rfloor -2}).\end{aligned}$$

The condition $A_1 \cap \cdots \cap A_i$ implies that there are $k \ge i$ steps in the phase in which bit position $i+1$ is not flipped. In the remaining $s-k$ steps, the probability to select $x^{\mathrm{low}}$ is at most $8/n$, and the probability to mutate a specific bit position is $1/n$. Thus, for all $i \in I$,

$$\begin{aligned}\mathrm{Prob}(A_{i+1} \mid A_1 \cap \cdots \cap A_i) &\le 1-(1-8/n^2)^{s-k}\\ &\le 1-(1-8/n^2)^{s} \le 1-e^{-(1-\varepsilon)\ln n} \le 1-1/n^{1-\varepsilon}.\end{aligned}$$

Hence,

$$\mathrm{Prob}(S) \le (n/8)(1-1/n^{1-\varepsilon})^{n/8-1} \le e^{\ln(n/8)-(n/8-1)n^{\varepsilon-1}} = e^{-\Omega(n^{\varepsilon})}.$$

The sum of the last failure probability and the failure probability in the initial step is $e^{-\Omega(n^{\varepsilon})}$. □

The next step is to show that our results for the local SEMO (Theorem 4 and Theorem 5) carry over to the global SEMO. Our upper bound for the global SEMO (Theorem 6) is of the same quality as in the case of the local SEMO. Only the lower bound (Theorem 7) is slightly weaker. Yet, it still guarantees time $\Omega(n^2 \log n)$ with high probability.

**Lemma 2.** *For the mutation operator of the global SEMO and $0 \le \varepsilon \le 1$ a constant, the following holds. The probability that any of at most polynomially many (in $n$) mutation steps flips at least $n^{\varepsilon}$ bits is $n^{-\Omega(n^{\varepsilon})}$.*

*Proof.* The probability that a single mutation step flips at least $k \ge 0$ bits is at most

$$\binom{n}{k}\left(\frac{1}{n}\right)^k \le \frac{1}{k!} \le \left(\frac{e}{k}\right)^k = e^{-k\ln k + k}.$$

In a series of polynomially many steps, say $cn^d$ where $c \ge 0$ and $d \ge 0$ are constants, the probability that any step flips at least $k$ bits is upper bounded by the sum of the probabilities for each step. The latter is at most $e^{-k\ln k+k} \cdot cn^d = e^{-k\ln k+k+d\ln n+\ln c}$. For $k = n^{\varepsilon}$, the last expression is $e^{-\Omega(n^{\varepsilon}\ln n)} = n^{-\Omega(n^{\varepsilon})}$. □

**Theorem 6.** *For any polynomial $p(n) > 0$, the runtime of the global SEMO for* MOCO *is $O(n^2 \log n)$ with a probability of at least $1-1/p(n)$.*

*Proof.* This proof has the same structure as the proof of Theorem 4. We show that the success probability of each epoch is essentially unchanged.

We show that $n^2/4 + n$ steps produce a search point with less than $n/2$ ones with probability $1 - e^{-\Omega(n)}$. We call a step *relevant* if either $x^{\text{low}}$ or $x^{\text{high}}$ is selected for mutation. The expected number of relevant steps is at least $n$ and, by Chernoff bounds, the probability of at least $(21/32)n$ relevant steps is $1 - e^{-\Omega(n)}$. We consider a phase starting with the step right after the initial step up to the $\lfloor (21/32)n \rfloor$th relevant step. Since the selection mechanism selects individuals uniformly at random, each relevant step selects $x^{\text{low}}$ with probability $1/2$. By Chernoff bounds, at least $(9/32)n$ relevant steps select $x^{\text{low}}$ and at most $(12/32)n$ relevant steps select $x^{\text{high}}$ with a probability $1 - e^{-\Omega(n)}$. Since $x^{\text{low}}$ has at least $n/2$ ones, every step mutating $x^{\text{low}}$ decreases the number of ones with a probability of at least $(n/2)(1/n)^1(1-1/n)^{n-1} \geq 1/(2e) \geq 1/6$. Thus, $\|x^{\text{low}}\|$ decreases an expected number of at least $(1/6)(9/32)n = (3/2)(n/32)$ times by at least 1. By Chernoff bounds, the sum of the decreases is at least $(n/32)$ with a probability $1 - e^{-\Omega(n)}$. The initial search point has less than $(17/32)n$ ones with a probability $1 - e^{-\Omega(n)}$. We show that $\|x^{\text{low}}\|$ is unlikely to increase in any step of the phase. This could only happen if $\|x^{\text{high}}\| > (31/32)n$. By Lemma 2, our phase of $O(n^2)$ steps has no step flipping at least $(1/64)n$ bits with a probability of at least $1 - n^{-\Omega(n)}$. As we can exclude that $\|x^{\text{high}}\|$ increases by at least $(1/64)n$ in a step, there must be a point of time where $\|x^{\text{high}}\|$ is at least $\lfloor (3/4 + 1/64)n \rfloor$ and at most $\lfloor (3/4 + 2/64)n \rfloor$. Since there can be at most one individual from Region IV in $A$, $\|x^{\text{high}}\|$ can only increase if $x^{\text{high}}$ is selected for mutation, i.e., in a relevant step. We pessimistically assume that $\|x^{\text{high}}\|$ initially is $\lfloor (3/4 + 2/64)n \rfloor = \lfloor (25/32)n \rfloor$. Since the number of zeroes in $x^{\text{high}}$ is at most $n/4$, in at most $(12/32)n$ relevant steps selecting $x^{\text{high}}$, there are at most $(3/32)n^2$ chances to flip a zero into a one. The expected number of flipping zeroes is at most $(3/32)n$. By Chernoff bounds, at most $(5/32)n$ new ones are created with a probability $1 - e^{-\Omega(n)}$. The phase is finished successfully with a probability of $1 - e^{-\Omega(n)}$.

In the second epoch, the probability of a successful step is lower bounded by

$$\frac{1}{n/4+1}\binom{n/4}{1}\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{n-1} \;\geq\; \frac{1}{e(n+4)}.$$

By Chernoff bounds, the next $O(n^2)$ steps produce a point with at most $n/4$ ones with a probability $1 - e^{-\Omega(n)}$.

For the last epoch, let $x^\ell$ denote the first individual with $\ell \leq n/4$ ones. The probability to create the next individual $x^i$ in $x^{\ell+1}, \ldots, x^{n/4}$ in the next step is at least

$$\frac{1}{n/4+1}\binom{n-i}{1}\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{n-1} \;\geq\; \frac{3}{e(n+4)}.$$

By Chernoff bounds, the next $O(n^2)$ steps produce all $x^{\ell+1}, \ldots, x^{n/4}$ with probability $1 - e^{-\Omega(n)}$. Now we consider the points $x^{\ell-1}, \ldots, x^0$. Initially, we set $j = \ell$ and call a step *relevant* if $x^j$ is selected for mutation. If solely a one flips in a relevant step, we decrease $j$ by 1 regardless of the offspring $x'$ being accepted or not. If $x'$ is not accepted, there is already an individual $x^{j-1}$ in $A$. (Note that $x^j$ is not necessarily the individual with the least number of ones.) If the offspring

$x'$ is not accepted, we identify each one in $x'$ with a one in $x^{j-1}$. This way, we can track each one of the original point $x^\ell$ in the points $x^j$, $j < \ell$, although the positions of the ones vary. Now it is clear that each one of the original point $x^\ell$ can decrease $j$ by 1. The probability that the next step is relevant and flips solely a specific one is at least

$$\frac{1}{n/4+1}\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{n-1} \geq \frac{4}{en^2+4en} =: p.$$

The probability that in any of $r := \lceil((d+3)/p)\ln n\rceil = O(n^2 \log n)$ steps a specific one does not decrease $j$ is at most $(1-p)^r \leq n^{-(d+3)}$. The probabilities that there is a one among the at most $n/4$ ones of $x^\ell$ that does not decrease $j$ is upper bounded by the sum of the corresponding probabilities $\sum_{i=1}^{n/4} n^{-(d+3)} \leq (1/4)n^{-(d+2)}$. Taking into account the failure probability for $x^{\ell+1}, \ldots, x^{n/4}$, $O(n^2 \log n)$ steps finish the last epoch with a probability $1 - n^{-(d+2)}$. □

**Theorem 7.** *The runtime of the global SEMO for* MOCO *is* $\Omega(n^2 \log n)$ *with a probability* $1 - O(1/\sqrt{n})$.

*Proof.* Analogously to the proof of Theorem 5 we show that the exploration of Region I takes $\Omega(n^2 \log n)$ steps in a typical run. Our first aim is to show the following property. With probability $1-O(1/\sqrt{n})$, there is a point of time where the number of individuals in Region I is at least $n/300$ and $n^{1/8} \leq \|x^{\text{low}}\| \leq 2n^{1/8}$, or the runtime is at least $n^2 \log n$ anyway.

The initial point has more than $n/4$ ones with a probability $1 - e^{-\Omega(n)}$. Then we wait for the event that a point with at most $2n^{1/8}$ ones is created, i.e., $\|x^{\text{low}}\| \leq 2n^{1/8}$. If this takes at least $n^2 \log n$ steps, we are done. Otherwise we consider a phase of $\lfloor n^2/50\rfloor$ steps subsequent to the step that created the first individual with at most $2n^{1/8}$ ones. At any time, let $g$ be the smallest number in $\{\|x^{\text{low}}\|, \ldots, n/4\}$ such that $x^{g-1}$ is in $A$ but $x^g$ is still missing. If that number $g$ does not exist, the number of individuals in Region I is at least $n/4 - 2n^{1/8} \geq n/300$ for $n$ large enough. In this case, we consider the phase finished. In the phase, a step is called *relevant* if it selects either $x^{g-1}$ or any of the two individuals with the two least numbers of ones, but at most $2n^{1/8}$ ones. At least the individual $x^{\text{low}}$ fulfills these conditions. Hence, the probability of a relevant step is at least $1/(n/4+1) \geq 15/(4n)$ for $n$ large enough. The expected number of relevant steps in $\lfloor n^2/50\rfloor$ steps is at least $(3/40)n-1$, and, by Chernoff bounds, there are at least $n/20$ relevant steps with a probability $1-e^{-\Omega(n)}$. We finish the phase after the $\lfloor n/20\rfloor$th relevant step if this step is prior to the $\lfloor n^2/50\rfloor$th step. A relevant step selects $x^{g-1}$ for mutation with a probability of at least 1/3. Whenever $x^{g-1}$ is selected, the probability to create $x^g$ by mutation is at least $(3n/4)(1/n)(1-1/n)^{n-1} \geq 3/(4e) \geq 1/4$ because $x^{g-1}$ has at least $3n/4$ zeroes. Thus, we expect at least $n/240 - 1/12$ new individuals in Region I in $\lfloor n/20\rfloor$ relevant steps. By Chernoff bounds, with probability $1-e^{-\Omega(n)}$ there are at least $n/300$ (new) individuals in Region I after at most $n^2/50$ steps of the phase.

We still have to show that no individual with less than $n^{1/8}$ ones has been created – neither in the phase nor in any of at most $n^2 \log n$ steps before the phase. A step is called *bad* if it either selects an individual with at most $3n^{1/8}$ ones and decreases the number of ones by at least 3 or if it selects an individual with more than $3n^{1/8}$ ones and flips more than $n^{1/8}$ bits. By Lemma 2, the probability that the second event occurs in the steps between the initial step and the end of the phase is at most $n^{-\Omega(n^{1/8})}$. The first event can only happen in the phase. We estimate the probability that a mutation step decreases the number of ones of an individual with at most $i \le \alpha n^\varepsilon$ ones by at least $k \le i$, $0 < \varepsilon \le 1$ and $\alpha > 0$ a constant. Letting $j$ denote the number of flipping ones, it is at most

$$\sum_{k\le j\le i} \binom{i}{j} \left(\frac{1}{n}\right)^j \le \sum_{k\le j\le i} \frac{i^j}{n^j} \le \sum_{k\le j\le i} \alpha^j n^{(\varepsilon-1)j} \le \alpha^k n^{(\varepsilon-1)k} \sum_{j\ge 0} \left(\frac{\alpha}{n^{(1-\varepsilon)}}\right)^j \le 2\alpha^k n^{(\varepsilon-1)k}, \quad (5)$$

for $n$ large enough. Hence, given that a step has selected an individual with at most $3n^{1/8}$ ones, it decreases the number of ones in the selected individual by at least 3 with a probability of at most $54n^{-21/8}$. Thus, the probability that the second event occurs in the phase is at most $54n^{-21/8} \cdot n^2/50 = O(n^{-5/8})$, and there is no bad step prior to the end of the phase with probability $1 - O(n^{-5/8})$. Now we work under the condition that there is no bad step. That means, individuals with at most $2n^{1/8}$ ones are offspring of individuals with at most $3n^{1/8}$ ones, and if an individual with at most $3n^{1/8}$ ones is selected, $\|x^{\text{low}}\|$ decreases by at most 2. In the phase, the value $\|x^{\text{low}}\|$ can only decrease if one of the two individuals with the smallest numbers of ones is selected, i. e., in a relevant step. Hence, the individual with at most $2n^{1/8}$ ones that is created first has at least $2n^{1/8} - 1$ ones. At least $\lfloor (1/2)n^{1/8} \rfloor$ relevant steps that decrease $\|x^{\text{low}}\|$ (by at most 2) are required in the phase to create an individual with at most $n^{1/8}$ ones. We pessimistically assume that each relevant step selects one of the two individuals with the least numbers of ones. For a step that has selected an individual with at most $3n^{1/8}$ ones, let $A$ be the event that it decreases the number of ones by at least 1, and let $B$ be the event that it is not bad, i. e., the number of ones decreases by at most 2. We estimate the conditional probability that the number of ones decreases by at least 1 given that the step is not bad. By (5), $\text{Prob}(A) \le 6/n^{7/8}$ and $\text{Prob}(B) \ge 1 - 54/n^{21/8}$. Hence, $\text{Prob}(A \mid B) \le \text{Prob}(A)/\text{Prob}(B) \le 7/n^{7/8}$ for $n$ large enough. Thus, $\lfloor n/20 \rfloor$ relevant steps in a good phase have an expected number of at most $n/20 \cdot 7/n^{7/8} = (7/20)n^{1/8}$ such decreasing steps and, by Chernoff bounds, with probability $1 - e^{-\Omega(n^{1/8})}$, this number is at most $\lfloor (1/2)n^{1/8} \rfloor$. No individual with less than $n^{1/8}$ ones has been created when the phase is over with probability $1 - O(n^{-5/8})$. Taking also into account the failure probability in the initial step and the probability that the population in Region I does not grow to size $n/300$, the first aim is reached with a probability of at least $1 - O(1/\sqrt{n})$.

Now we assume that a situation in accordance with the second condition of our first aim has been reached. That means, $n^{1/8} \leq \|x^{\text{low}}\| \leq 2n^{1/8}$, and there are at least $n/300$ individuals from Region I in $A$. The optimization process is not finished before $x^1$ has been created. We consider a new phase consisting of the next $s = \lfloor (1/16)(n^2/300 - 1) \ln n \rfloor = \Omega(n^2 \log n)$ steps and claim that this phase will not produce $x^1$ with probability $1 - O(1/\sqrt{n})$. This is our second aim. First we show that a newly created individual $x^{\text{low}}$ is a direct offspring of the former $x^{\text{low}}$ with high probability. We call the phase *clean* if this condition is not violated. In other words, in a clean phase, no individual with strictly more than $\|x^{\text{low}}\|$ ones produces an offspring with strictly less than $\|x^{\text{low}}\|$ ones. This property can only be violated if at least $n^{1/8}$ bits flip in an individual with more than $3n^{1/8}$ ones or if the number of ones decreases by at least 2 in an individual with at most $3n^{1/8}$ ones excluding $x^{\text{low}}$. The probability that the first event occurs in the phase is $n^{-\Omega(n^{1/8})}$ by Lemma 2. Each step selects an individual with at most $3n^{1/8}$ ones with a probability of at most $3n^{1/8}/(n/300) = 900n^{-7/8}$, and, by (5), the probability to decrease the number of ones by at least 2 is at most $18n^{-14/8}$. The probability that the second event occurs in the phase is at most $900n^{-7/8} \cdot 18n^{-14/8} \cdot (n^2/(16 \cdot 300)) \ln n \leq 4n^{-5/8} \ln n$. Thus, the phase is clean with probability $1 - O(1/\sqrt{n})$. In the following, we assume that none of the two events occurs in the phase, i.e., the phase is clean. This condition does neither affect the probability to select $x^{\text{low}}$ for mutation nor the probability that a bit flips when $x^{\text{low}}$ has been selected. Now we argue analogously to the proof of Theorem 5. Let $I$ be a set of $\lfloor n^{1/8} \rfloor$ bit positions of ones in $x^{\text{low}}$ by the time that the phase begins. W.l.o.g., let $I = \{1, \ldots, \lfloor n^{1/8} \rfloor\}$. Let $A_i$ denote the event that the bit at position $i$ flips at least once in the steps that select $x^{\text{low}}$. Then the phase can only create $x^1$ if at least $\lfloor n^{1/8} \rfloor - 1$ of these events $A_i$ happen. Let $S$ be the event that $x^1$ is created in the phase. Then $\text{Prob}(S)$ is at most

$$\begin{aligned} &\text{Prob}\left((A_2 \cap \cdots \cap A_{\lfloor n^{1/8} \rfloor}) \cup \cdots \cup (A_1 \cap \cdots \cap A_{\lfloor n^{1/8} \rfloor - 1})\right) \\ &\leq \lfloor n^{1/8} \rfloor \, \text{Prob}(A_1 \cap \cdots \cap A_{\lfloor n^{1/8} \rfloor - 1}) \end{aligned}$$

since the union of events is upper bounded by the sum of the probabilities of each event, and the events $A_i$ have equal probabilities. In the case of independent bit flips, the events $A_i$ are independent events. Therefore,

$$\text{Prob}(S) \leq \lfloor n^{1/8} \rfloor \, \text{Prob}(A_1) \cdots \text{Prob}(A_{\lfloor n^{1/8} \rfloor - 1}) = \lfloor n^{1/8} \rfloor \left(\text{Prob}(A_1)\right)^{\lfloor n^{1/8} \rfloor - 1}.$$

The probability that a bit at a specific position $i \in I$ of $x^{\text{low}}$ flips at least once in the phase can be estimated in the following way. The individual $x^{\text{low}}$ is selected with probability at most $1/(n/300)$ and the bit flips with probability $1/n$. Hence,

$$\text{Prob}(A_i) \leq 1 - (1 - 300/n^2)^{(1/16)(n^2/300 - 1) \ln n} \leq 1 - e^{-(1/16) \ln n} = 1 - n^{-1/16},$$

for all $i \in I$. Now we have

$$\text{Prob}(S) \leq n^{1/8} (1 - n^{-1/16})^{\lfloor n^{1/8} \rfloor - 1} \leq e^{(1/8) \ln n - ((\lfloor n^{1/8} \rfloor - 1)/n^{1/16})} = e^{-\Omega(n^{1/16})}.$$

Hence, the second phase of $\Omega(n^2 \log n)$ steps fails to produce $x^1$ with probability $1 - O(1/\sqrt{n})$. □

## 6 A Test Function

The functions considered in this section are inspired by the well-known function $x \mapsto (x^2, (x-2)^2)$. The latter often serves as a test function for algorithms that work in the continuous decision space $\mathbb{R}$ (e.g., Srinivas and Deb (1994)). We adapt this function to the Boolean decision space in two different ways. The first variant is based on the CountingOnes function uses a kind of unary encoding of integer numbers, the second one the standard binary encoding.

**Definition 10.** *For $x = x_{n-1}, \ldots, x_0 \in \{0,1\}^n$, let $\|x\| = \sum_{0 \le i \le n-1} x_i$ denote the number of ones in $x$ and $\mathrm{BV}(x) = \sum_{0 \le i \le n-1} x_i 2^i$ the binary value of $x$. The functions $f_{a,b}$, $0 \le a < b \le n$, and $g_{a,b}$, $0 \le a < b \le 2^n - 1$, are defined by*

$$f_{a,b}(x) = \big((\|x\| - a)^2, (\|x\| - b)^2\big),$$
$$g_{a,b}(x) = \big((\mathrm{BV}(x) - a)^2, (\mathrm{BV}(x) - b)^2\big).$$

For both functions, the goal is to minimize the two objectives. We adapt the basic definitions to the case of minimization. In particular, we redefine (1) by $y \preceq z :\Leftrightarrow \forall i\colon y_i \ge z_i$.

**Proposition 5.** *The Pareto set and Pareto front of $f_{a,b}$ are*

$$X^* = \{x \mid a \le \|x\| \le b\} \quad \textit{resp.} \quad F^* = \{(i^2, (b-a-i)^2) \mid 0 \le i \le b-a\}.$$

*Proof.* Any point $x$ with $\|x\| < a$ ($\|x\| > b$) is not Pareto optimal since the value of both objectives decreases as $\|x\|$ increases (decreases) by 1. Consider a point $z$, $a \le \|z\| \le b$. We show that $z$ is not dominated by any point $w$, i.e., $z$ is Pareto optimal. If $\|z\| = \|w\|$ then $z =_{f_{a,b}} w$. If $\|z\| < \|w\|$ then $(\|z\| - a)^2 < (\|w\| - a)^2$ holds and implies $z \not\preceq_{f_{a,b}} w$. If $\|z\| > \|w\|$ then $(\|z\| - b)^2 < (\|w\| - b)^2$ holds and implies $z \not\preceq_{f_{a,b}} w$. For all $z$ with $\|z\| = a + i$, the corresponding objective vector is $(((a+i) - a)^2, ((a+i) - b)^2)$. □

**Proposition 6.** *The Pareto set and Pareto front of $g_{a,b}$ are*

$$X^* = \{x \mid a \le \mathrm{BV}(x) \le b\} \quad \textit{resp.} \quad F^* = \{(i^2, (b-a-i)^2) \mid 0 \le i \le b-a\},$$

*and $|X^*| = |F^*|$.*

*Proof.* Can be carried out analogously to the proof of Proposition 5. □

**Theorem 8.** *The expected runtime of the local and global SEMO for $f_{a,b}$ is*

$$O\big(n \log n + n(b-a)\log(b-a)\big).$$

Note that $b - a + 1$ is the cardinality of the Pareto front. If $a$ and $b$ are constants (as in $x \mapsto (x^2, (x-2)^2)$), the expected runtime is $O(n \log n)$.

*Proof.* In this proof, all bounds are derived for the global SEMO but they also hold for the local SEMO because we merely consider mutation steps where a single bit flips. We partition the process into two epochs. The first epoch is the time before the first search point in the Pareto set is produced and the second epoch is the remaining time until the image of the population is exactly the Pareto front.

For the first epoch, note that the population size $|A|$ is bounded by 2. At any time, there is at most one search point $x^{\text{low}}$ such that $\|x^{\text{low}}\| < a$ because any other point with this property would either dominate $x^{\text{low}}$ or be dominated by $x^{\text{low}}$. For the same reason, there is at most one search point $x^{\text{high}}$ such that $\|x^{\text{high}}\| > b$. Let $x \in A$ in the first epoch. From a local point of view, the aim for $x$ is to increase (decrease) the number of ones if $\|x\| < a$ ($\|x\| > b$) . If we consider only $x$, the scenario is similar to the situation where the (1+1) EA optimizes the function OneMax (ZeroMax) (Droste, Jansen, and Wegener (2002)). At any time, let $x$ be the individual in the population such that $d = |\|x\| - a|$ takes the smaller value; ties broken arbitrarily. The $d$-value is non-increasing in the first epoch since $d^2$ is the first objective of individual $x$, and $x$ can only be dominated by a new individual with a $d$-value that is not larger. Notice that we can always specify $d$ bits of $x$ such that flipping these bits would decrease the $d$-value to 0. The probability that the $d$-value decreases in the next step is lower bounded by the probability that the next step chooses $x$ for mutation and flips solely one bit out of $d$ specified bits in $x$. The latter probability is at least $(1/2)d(1/n)(1-1/n)^{n-1} \geq d/(2en)$. The expected time until the first Pareto optimal search point is produced is at most

$$\sum_{n \geq d \geq 1} \frac{2en}{d} = 2enH_n = O(n \log n),$$

where $H_n$ denotes the $n$th harmonic number.

We argue that for each point of time $t < T_{f_{a,b}}$ in the second epoch, the size of the population is at most $b-a$. If there are only Pareto optimal search points in the population, this property follows from Proposition 5 as $A \neq X^*$ in the second epoch. Now consider the case that there are non-Pareto optimal points in $A$. We have already seen that there are at most two individuals $x^{\text{low}}, x^{\text{high}} \in \{0,1\}^n - X^*$ in the population, where $\|x^{\text{low}}\| < a$ and $\|x^{\text{high}}\| > b$. If only one of them exists, say $x^{\text{low}}$ (implying $a \geq 1$), then we have to show that there are strictly less than $b-a$ points of the Pareto set in the population. As the minimum value of the first objective of $x^{\text{low}}$ is 1, we can exclude at least all points in $X^*$ whose first objective takes a value of 0 or 1, i.e., all points $x \in X^*$ with $\|x\| \in \{a, a+1\}$. The remaining $b-a-1$ points in the Pareto front are represented by at most that many individuals. If only $x^{\text{high}}$ exists (implying $b < n$), we can exclude all points $x \in X^*$ with $\|x\| \in \{b-1, b\}$ using analogous arguments. If both $x^{\text{low}}$ and $x^{\text{high}}$ exist, we can exclude all points in $X^*$ with $\|x\| \in \{a, a+1, b-1, b\}$. In the last case, there are at most $b-a-3$ Pareto optimal points in the population plus $x^{\text{low}}$ and $x^{\text{high}}$.

Now we estimate the probability $p_i$, $a \le i \le b$, that a search point with $i$ ones is created in the next step, given that there is already a search point $x$ in the population with $i-1$ or $i+1$ ones. The algorithm selects $x$ for mutation with a probability of at least $1/(b-a)$. The probability that the mutation step creates a string with exactly $i$ ones from a string with $i-1$ ones is at least $(n-i+1)(1/n)(1-1/n)^{n-1}$; the probability that the mutation step creates such a string from a string with $i+1$ ones is at least $(i+1)(1/n)(1-1/n)^{n-1}$. We only underestimate the probability $p_i$ if we use the bounds

$$p_i \ge \begin{cases} \frac{n-i+1}{b-a}\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} \ge \frac{n-i+1}{(b-a)ne} & \text{if } i > n/2, \\ \frac{i+1}{b-a}\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} \ge \frac{i+1}{(b-a)ne} & \text{if } i \le n/2. \end{cases}$$

Let $T_i$ denote the waiting time until a string with $i$ ones is created given that a string with $i-1$ or $i+1$ ones has been created before. Then $E(T_i)$ is at most $1/p_i$. Let $j$ be the number of ones of the first point in the Pareto set created by the algorithm. The expected duration of the second epoch is at most

$$E(T_{j-1}) + \cdots + E(T_a) + E(T_{j+1}) + \cdots + E(T_b).$$

Our bounds for $E(T_0)$ and $E(T_n)$ are the largest, namely $(b-a)ne/1$. For $E(T_1)$ and $E(T_{n-1})$, they are $(b-a)ne/2$ and so on. Hence, the last sum is upper bounded by the sum of the $b-a$ largest bounds. The latter is at most

$$\begin{aligned} &(b-a)ne \cdot 2\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{\lceil (b-a)/2\rceil}\right) \\ &= 2en(b-a)H_{\lceil (b-a)/2\rceil} \;=\; O\big(n(b-a)\log(b-a)+n\big). \end{aligned}$$ □

If we switch to the standard bit representation of integer numbers, the function turns intractable for the local SEMO. The reason is that there can be large Hamming cliffs which are not negotiable for the local mutation operator. Also for the global SEMO the function can become much harder. We show that, for each $d \in \{2, \ldots, n\}$, one can choose $a$ and $b$ such that the expected runtime of the global SEMO is $\Theta(n^d)$.

**Theorem 9.** *Let $a = 2^k - 1$, $b = 2^k$, and $1 \le k \le n-1$. The expected runtime of the global SEMO for $g_{a,b}$ is $\Theta(n^{k+1})$. Moreover, for $0 < c < 1$, the runtime is at least $n^{c(k+1)}$ with a probability of at least $1 - \frac{1}{n^{(1-c)(k+1)}}$ and at most $n^{k+2}$ with a probability $1 - e^{-\Omega(n)}$.*

*Proof.* According to Proposition 6, the Pareto front is $F^* = \{(0,1),(1,0)\}$, and the Pareto set is $X^* = \{\mathrm{BV}^{-1}(a), \mathrm{BV}^{-1}(b)\} = \{0^{n-k}1^k, 0^{n-k-1}10^k\}$ where $\mathrm{BV}^{-1}(i)$ denotes the bit representation of a non-negative integer $i$. We partition the run of the algorithm into two consecutive epochs. The first epoch lasts until the first Pareto optimal search point $\mathrm{BV}^{-1}(a)$ or $\mathrm{BV}^{-1}(b)$ is created. For the first epoch, we prove only a weak upper bound because the second epoch dominates the runtime; however, we must take care that our bound holds with a probability $1 - e^{-\Omega(n)}$.

Note that the population size $|A|$ is bounded by 2 in the first epoch. There is at most one search point $x^{\text{low}}$ such that $\text{BV}(x^{\text{low}}) < a$ because any other point with this property would either dominate $x^{\text{low}}$ or be dominated by $x^{\text{low}}$. For the same reason, there is at most one search point $x^{\text{high}}$ such that $\text{BV}(x^{\text{high}}) > b$. We subdivide the first epoch into two subepochs such that the first subepoch lasts until an individual $x$ with $\text{BV}(x) < 2^{k+1}$ is created. Clearly, the population is $A = \{x^{\text{high}}\}$ in the first subepoch. If solely the leftmost one of $x^{\text{high}}$ flips, $\text{BV}(x^{\text{high}})$ is at least halved. We call this event a success in the first subepoch. The probability that a step is a success is $(1/n)(1-1/n)^{n-1} \geq 1/(ne)$. As the initial value of $\text{BV}(x^{\text{high}})$ is at most $2^n - 1$, a number of $n-(k+1)$ successes are sufficient in the first subepoch. In the second subepoch, the binary value of each individual in $A$ is less than $2^{k+1}$, i.e., all prefix bits corresponding to the weights $2^{n-1}, \ldots, 2^{k+1}$ are 0-bits. For $x \in A$, let $d(x) := \min\{|\text{BV}(x) - a|, |\text{BV}(x) - b|\}$, and let $d(A) = \min\{d(x), x \in A\}$, i.e., $d(A)$ is the smallest distance from a point in $A$ to a point in the Pareto set in terms of binary values. At any time, let $x \in A$ be the point with the smaller $d(x)$-value; ties broken arbitrarily. Note that if $x$ is removed from $A$, a new individual with a $d$-value that is not larger enters the population at the same time. Consequently, the $d(A)$-value only decreases with time. If $\text{BV}(x) > b$ then $x_k = 1$ and $d(A) = \text{BV}(x) - b = \sum_{0 \leq i \leq k-1} x_i 2^i$. Flipping solely the leftmost 1-bit in the suffix $x_{k-1}, \ldots, x_0$ reduces the $d(A)$-value at least by a factor of 1/2. If $\text{BV}(x) < a$ then $x_k = 0$ and $d(A) = a - \text{BV}(x) = \sum_{0 \leq i \leq k-1} (1 - x_i) 2^i$. Flipping solely the leftmost 0-bit in the suffix $x_{k-1}, \ldots, x_0$ reduces the $d(A)$-value at least by a factor of 1/2. The algorithm selects $x$ for mutation with a probability of at least 1/2. Hence, the next step decreases the $d(A)$-value at least by a factor of 1/2 with a probability of at least $(1/2)(1/n)(1-1/n)^{n-1} \geq 1/(2en)$. We call this event a success in the second subepoch. A number of $k$ successes are sufficient for the second subepoch, and less than $n$ successes are sufficient for the first epoch. In a sequence of $12n^2$ steps, the expected number of successes is at least $2n$ and, by Chernoff bounds, the probability of less than $n$ successes is $e^{-\Omega(n)}$.

By the time that the second epoch starts, the algorithm has found either $\text{BV}^{-1}(a)$ or $\text{BV}^{-1}(b)$ first. The point $\text{BV}^{-1}(a)$ ($\text{BV}^{-1}(b)$) dominates all other points in the decision space except $\text{BV}^{-1}(b)$ ($\text{BV}^{-1}(a)$). Therefore, the population is $\{\text{BV}^{-1}(a)\}$ or $\{\text{BV}^{-1}(b)\}$, and no offspring except $\text{BV}^{-1}(b)$ resp. $\text{BV}^{-1}(a)$ will be accepted. Only a mutation step flipping solely the $k+1$ rightmost bits corresponding to the weights $2^k, \ldots, 2^0$ will be accepted. The corresponding probability is $(1/n)^{k+1}(1-1/n)^{n-(k+1)}$. It is upper and lower bounded by $1/n^{k+1}$ and $1/(en^{k+1})$, respectively. Thus, the expected waiting time for this event is upper and lower bounded by the expectations of random variables following the geometric distribution with parameter $1/(en^{k+1})$ and $1/(n^{k+1})$, respectively. Hence, the expected runtime (for both epochs) is $\Theta(n^{k+1})$.

The probability that a number of steps in the second epoch succeeds in producing the second Pareto optimal point is upper bounded by the sum of the success probabilities in each step. Hence, the probability that the first $n^{c(k+1)}$ steps in the second epoch are not successful is lower bounded by

$$1 - \frac{1}{n^{(k+1)}} n^{c(k+1)} = 1 - \frac{1}{n^{(1-c)(k+1)}}.$$

Remember that the first epoch is finished after $12n^{k+1}$ steps with an overwhelming probability $1 - e^{-\Omega(n)}$. The first $n^{k+2} - 12n^{k+1}$ steps in the second epoch succeed in finding the second Pareto optimum with a probability of at least

$$1 - \left(1 - \frac{1}{en^{k+1}}\right)^{n^{k+2}-12n^{k+1}} \geq 1 - e^{-\Omega(n)}. \qquad \square$$

Although one of the two Pareto optima is found quickly by the algorithm (almost surely in time $O(n^2)$), a large Hamming distance to the second Pareto optimum ensures a large (expected) runtime. For $k = \Theta(n)$, the runtime is $n^{\Theta(n)}$ with a probability exponentially close to 1. Apparently, the global and local SEMO would not always find the same Pareto optimum first. Multiple runs could help to detect the entire Pareto set if each instance of the algorithm is halted after $12n^2$ steps and non-dominated solutions in the union of the final populations are computed.

## 7 Conclusion

The runtime of simple multi-objective evolutionary algorithms (MOEAs) can be analyzed. In the worst case, the global SEMO has an expected runtime of $O(n^n)$ that matches the expected worst-case runtime of simple EAs working in the scenario of single-objective optimization. The expected worst-case runtime is independent of the dimension of the objective space. Moreover, for each $d \in \{2, \ldots, n\}$, we have exhibited a bicriteria problem such that the expected runtime of the global SEMO is $\Theta(n^d)$. Explicit bounds on the expected runtime for simple objective functions can be derived, e. g., for the problem LOTZ (leading ones trailing zeroes). In many situations, bounds on the runtime that hold with high probability can be derived. Applied to the problem MOCO (multi-objective counting ones), the global and local SEMO have very large resp. non-finite expected runtimes but the runtimes are $\Theta(n^2 \log n)$ with high probability.

## Acknowledgments

I am grateful to Marco Laumanns for discussions on multi-objective EAs and their analysis. I thank Ingo Wegener and Carsten Witt for several suggestions for improvement.

## Bibliography

Beyer, H.-G., Schwefel, H.-P., and Wegener, I. (2002). How to analyse evolutionary algorithms. *Theoretical Computer Science* 287, 101–130.

Droste, S., Jansen, T., Tinnefeld, K., and Wegener, I. (2003). A new framework for the valuation of algorithms for black-box optimization. *Proc. of the 7th Foundations of Genetic Algorithms Workshop (FOGA 7)*, 253–270.

Droste, S., Jansen, T., and Wegener, I. (1998). On the optimization of unimodal functions with the (1+1) evolutionary algorithm. *Proc. of the 5th Conf. on Parallel Problem Solving from Nature (PPSN V),* LNCS 1498, 13–22.

Droste, S., Jansen, T., and Wegener, I. (2002). On the analysis of the (1+1) evolutionary algorithm. *Theoretical Computer Science* 276, 51–81.

Garnier, J., Kallel, L., and Schoenauer, M. (1999). Rigorous hitting times for binary mutations. *Evolutionary Computation* 7(2), 173–203.

Jägersküpper, J. (2003). Analysis of a simple evolutionary algorithm for minimization in Euclidian spaces. *Proc. of the 30th Internat. Colloq. on Automata, Languages, and Programming (ICALP 2003)*, LNCS 2719, 1068–1079.

Laumanns, M., Thiele, L., Zitzler, E., Welzl, E., and Deb, K. (2002). Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem. *Proc. of the 7th Internat. Conf. on Parallel Problem Solving From Nature (PPSN VII)*, LNCS 2439, 44–53.

Motwani, R. and Raghavan, P. (1995). *Randomized Algorithms*. Cambridge University Press.

Rudolph, G. (1998a). Evolutionary search for minimal elements in partially ordered finite sets. *Proc. of the 7th Annual Conf. on Evolutionary Programming*, 345–353.

Rudolph, G. (1998b). On a Multi-Objective Evolutionary Algorithm and Its Convergence to the Pareto Set. *Proc. of the 5th IEEE Conf. on Evolutionary Computation*, 511–516.

Rudolph, G. (2001). Evolutionary search under partially ordered fitness sets. *Proc. of the Internat. NAISO Congress on Information Science Innovations (ISI 2001)*, 818–822.

Rudolph, G. and Agapie, A. (2000). Convergence properties of some multi-objective evolutionary algorithms. *Proc. of the 2000 Congress on Evolutionary Computation (CEC 2000)*, 1010–1016.

Scharnow, J., Tinnefeld, K., and Wegener, I. (2002). Fitness landscapes based on sorting and shortest paths problems. *Proc. of the 7th Conf. on Parallel Problem Solving from Nature (PPSN VII),* LNCS 2439, 54–63.

Srinivas, N. and Deb, K. (1994). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation* 2(3), 221–248.

Thierens, D. (2003). Convergence Time Analysis for the Multi-objective Counting Ones Problem. *Proc. of the 2nd Internat. Conf. on Evolutionary Multi-Criterion Optimizat ion (EMO 2003)*, LNCS 2632, 355–364.

Wegener, I. (2001). Theoretical aspects of evolutionary algorithms. *Proc. of the 28th Internat. Colloq. on Automata, Languages, and Programming (ICALP 2001)*, LNCS 2076, 64–78.

Wegener, I. and Witt, C. (2003). On the optimization of monotone polynomials by the (1+1) EA and randomized local search. *Proc. of the Genetic and Evolutionary Computation Conf. (GECCO 2003)*, LNCS 2723, 622–633.