INITIALIZING MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS WITH THE OPTIMA OF EACH OBJECTIVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-objective evolutionary algorithms (MOEAs) are among the most popular approaches to solve multi-objective optimization problems. In this work, we revisit a classic idea in multi-objective optimization, which is surprisingly little used in MOEAs, namely to start the optimization from the optima of the different objectives. The little theoretical previous work studying this idea in MOEAs has shown significant speed-ups for the theory-based GSEMO algorithm, but only together with a strong and non-standard diversity mechanism. In this work, we show that the GSEMO also profits from the initialization with the single-objective optima without additional non-standard mechanisms. We prove that this algorithm optimizes the ONEMINMAX and COCZ benchmarks in $O(n^{4/3}\log^{4/3} n)$ function evaluations, considerably faster than the classic $O(n^2 \log n)$ runtime known from previous works. We extend this result to the NSGA-II, the most prominent MOEA in practice. Here this speed-up is obtained as soon as each single-objective optimum is contained once in the typically large initial population. These results constitute the first proof that initializing standard multi-objective evolutionary algorithms with the optima of the objectives can significantly improve their performance.

1 INTRODUCTION

Most real-world optimization problems contain several, often conflicting objectives. Various methods have been proposed to solve such multi-objective optimization problems, with evolutionary algorithms being one of the most prominent approaches. While the success of multi-objective evolutionary algorithms (MOEAs) is undoubted in practice (Coello et al., 2007), in recent years, their theoretical understanding also made huge progress, e.g., with the first mathematical runtime analyses of algorithms like the NSGA-II, NSGA-III, SMS-EMOA, SPEA2, or PAES (Zheng et al., 2022; Wietheger & Doerr, 2023; Bian et al., 2023; Ren et al., 2024; Opris, 2025) and with several deep insights into their working principles (Do et al., 2023; Dang et al., 2024; Li et al., 2025; Ren et al., 2026).

One topic that is not understood very well so far is how to select the initial population of a MOEA. The vast majority of practical and almost all theoretical works use initial populations composed of random individuals. This is surprising, since non-evolutionary algorithms for multi-objective optimization problems often start with the optima of the individual objectives (Das & Dennis, 1998; Paquete & Stützle, 2003; Halffmann et al., 2022). Hence, this would also be a natural way to initialize a population of a MOEA. We note that the optima of the individual objectives, being the solutions of single-objective optimization problems, are generally easy to compute compared to solving a multi-objective problem. Thus, the additional cost of such an initialization will often be negligible.

Among the empirical results for MOEAs (e.g., Gong et al., 2023; 2025a;b; Benecke & Mostaghim, 2025; Benecke, 2026), adding single-objective optima generally improves the optimization performance. However, such insights are very little supported by mathematical rigor. To the best of our knowledge, in theoretical works, the idea of initializing the population of a MOEA with optima of the objectives has been used only by Qian et al. (2013) and, to a minor extent, in the two subsequent works (Li et al., 2016; Bian et al., 2018). Also, the initialization considered by Qian et al. (2013) was not the focus of the paper. Instead, this initialization was used together with a diversity mechanism to show that one-point crossover can improve the performance of the GSEMO algorithm on the COCZ and LOTZ/WLPTNO benchmarks. As we argue in Section 3, it is not clear to what extent the runtime gains shown there stem from the initialization or rather from the strong diversity mechanism.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our Contribution. We take a fresh look at the idea of initializing the population of a MOEA with the optima of the individual objectives. Our first main result (Theorem 3) is that a standard crossover-based GSEMO, that is, without any diversity mechanisms, also profits considerably from the initialization with the optima of the individual objectives. We prove that this algorithm computes the Pareto fronts of the ONEMINMAX and COCZ benchmarks, of problem size n, in expected time $O(n^{4/3}\log^{4/3} n)$, considerably faster than the best runtime bound of $O(n^2 \log n)$ known for the (G)SEMO with random initialization (Laumanns et al., 2004; Giel & Lehre, 2010). We note that different from Qian et al. (2013), we use uniform crossover, which appears more natural for problems like ONEMINMAX and COCZ, in which the position of a bit in the string has no specific meaning.

Since the GSEMO is an algorithm mostly used in theoretical analyses (due its simplicity), we extend our result to the NSGA-II (Theorem 9), the most prominent MOEA in practice (with, e.g., more than 60,000 citations for the paper by Deb et al. (2002) proposing it), and show a runtime guarantee of $O(N^{4/3}\log^{4/3} n)$ function evaluations, where $N \geq 4(n+1)$ is the population size of the algorithm. Interestingly, for our result to hold, it suffices that the initial population (of size N) contains one optimum for each objective—all other initial solutions can still be chosen randomly.

Overall, this first focused theoretical discussion of the initialization of a MOEA with the optima of the objectives suggests that this natural idea can lead to considerable performance gains. In particular for algorithms with large initial populations, our strong belief is that adding one optimum for each objective to the standard initial population is unlikely to be harmful but has the strong potential to lead to considerable performance gains. Due to the complex population dynamics showing up in our mathematical analysis, we regard in this work only the two benchmark problems ONEMINMAX and COCZ and the two algorithms (G)SEMO and NSGA-II. However, from the working principle detected in our proofs, we are optimistic that the positive message of this work extends to other settings as well.

2 PRELIMINARIES

Let $\mathbb{N}$ denote the set of non-negative integers. For all $a, b \in \mathbb{N}$, we write $[a..b]$ for the (discrete) interval $[a,b] \cap \mathbb{N}$, and we write $[b]$ for $[1..b]$. Let $n \in \mathbb{N}_{\geq 1}$. We call each bit string $x \in \{0,1\}^n$ an *individual*, we write $|x|_1$ for the number of 1-bits in x and $|x|_0$ for the number of 0-bits, and for each $i \in [n]$, let $x[i]$ denote the entry of x at position i. For all individuals $x, y \in \{0,1\}^n$, we denote their *Hamming distance* by $H(x,y) = |\{i \in [n] \mid x[i] \neq y[i]\}|$.

2.1 BI-OBJECTIVE MAXIMIZATION

Let $n \in \mathbb{N}_{\geq 1}$. We call a function $f\colon \{0,1\}^n \to \mathbb{R}^2, x \mapsto (f_1(x), f_2(x))$ an *objective function*, where $f_1, f_2\colon \{0,1\}^n \to \mathbb{R}$ are called the *single objectives*. For the remainder of this article, we reserve n for the dimension of the domain of f, and all our big-O asymptotics are with respect to this n.

For all $x, y \in \{0,1\}^n$, we say y *weakly dominates* x, written as $y \succeq x$, if $f_1(y) \geq f_1(x)$ and $f_2(y) \geq f_2(x)$. We say y *strictly dominates* x, written as $y \succ x$, if $y \succeq x$ and $f(x) \neq f(y)$. If $x \not\succeq y$ and $y \not\succeq x$, then x and y are *incomparable*.

An individual x is called *Pareto-optimal* if it is not strictly dominated by any other individual. The set of objective values of all Pareto-optimal solutions is called the *Pareto front* $F(f)$ of f. If for a (multi)set $S \subseteq \{0,1\}^n$, we have $F(f) \subseteq \{f(x) \mid x \in S\}$, we say that S *witnesses* the Pareto front. The goal is to compute a (multi)set of individuals that witnesses the Pareto front.

2.2 THE ONEMINMAX AND COCZ BENCHMARKS

The ONEMINMAX benchmark (Giel & Lehre, 2010) is the most widely studied benchmark for MOEAs for discrete search spaces, defined by ONEMINMAX$\colon \{0,1\}^n \to [0..n]^2; x \mapsto (|x|_0, |x|_1)$.

ONEMINMAX has a Pareto front of size $n+1$, which is given by $F(\text{ONEMINMAX}) = \{(i, n-i) \mid i \in [0..n]\}$, and the single-objective global optima of ONEMINMAX are 0^n and 1^n. Since for all individuals x we have $|x|_0 = n - |x|_1$, each element of the Pareto front is uniquely identified by the number of 1-bits of a corresponding individual. Therefore, we represent elements or subsets of the Pareto front only by their second objective value. Accordingly, we write $F(\text{ONEMINMAX}) = [0..n]$.

Algorithm 1: The (G)SEMO (Laumanns et al., 2002; Giel, 2003) framework with uniform crossover, maximizing a given objective function $f\colon \{0,1\}^n \to \mathbb{R}^2$, given an initial population $P_0 \subseteq \{0,1\}^n$ of mutually incomparable individuals and a crossover rate $p_c \in [0,1]$.

```
set P ← P_0;
while termination criterion not met do
    with probability p_c do
        choose x, y ∈ P uniformly at random with replacement;
        set z to be the result of uniform crossover with parents x and y;
    else
        choose x ∈ P uniformly at random;
        set z to be the result of mutation with parent x;
    if ∀w ∈ P: w ⊁ z then set P ← {w ∈ P | w ⊁̸ z} ∪ {z};
return P;
```

The COCZ (CountingOnesCountingZeros) benchmark (Laumanns et al., 2002; 2004) is similar to ONEMINMAX for the second half of the individuals, but it contains a cooperative part in the first half. Hence, not all individuals are Pareto-optimal. Formally, letting for all $a, b \subseteq [n]$ with $a < b$ and all $x \in \{0,1\}^n$ the notation $x[a..b]$ denote the individual $(x[i])_{i \in [a..b]}$, we have $\mathrm{COCZ}\colon \{0,1\}^n \to [0..n]^2, x \mapsto (|x|_1, |x[1..\lfloor \frac{n}{2} \rfloor]|_1 + |x[\lfloor \frac{n}{2} \rfloor + 1\,..\,n]|_0)$.

The Pareto front of COCZ is $F(\mathrm{COCZ}) = \{(\lfloor \frac{n}{2} \rfloor + i, n - i) \mid i \in [0..\lceil \frac{n}{2} \rceil]\}$ and thus has size $\lceil \frac{n}{2} \rceil + 1$. The single-objective global optima are 1^n and $1^{\lfloor n/2 \rfloor}0^{\lceil n/2 \rceil}$. Similar to ONEMINMAX, each element of the Pareto front is uniquely identified by the number of 1-bits in the second half of a corresponding individual. Thus, we represent elements or subsets of the Pareto front only by their first objective minus $\lfloor \frac{n}{2} \rfloor$. Accordingly, we write $F(\mathrm{COCZ}) = [0..\lceil \frac{n}{2} \rceil]$.

2.3 THE SEMO AND GSEMO MULTI-OBJECTIVE OPTIMIZERS

We consider the well-known *simple evolutionary multi-objective optimizer* (SEMO; Algorithm 1) (Laumanns et al., 2002) and its variant *global SEMO* (GSEMO) (Giel, 2003), which creates offspring in a slightly different way that allows for global search. Both algorithms feature a dynamically changing set P of individuals (the *population*) of pairwise non-dominated individuals, which it updates iteratively. Starting with a given initial population, the (G)SEMO produces in each iteration a new individual z (the *offspring*). If z is not strictly dominated by any element in P, it is added to P, and all other individuals weakly dominated by z are removed from P.

Offspring is produced based on existing individuals from P (the *parents*), always chosen uniformly at random (with replacement). The (G)SEMO selects, based on a given probability $p_c \in (0,1)$ (the *crossover rate*), two parents with probability p_c, and it selects one parent with probability $1 - p_c$. In case of two parents, the operation is called *crossover*, in the case of a single parent, *mutation*.

We consider *uniform crossover*, which creates offspring $z \in \{0,1\}^n$ from two parents $x, y \in \{0,1\}^n$ by selecting for each position uniformly from the respective bits of the parents. That is, for all $i \in [n]$, we have with probability $\frac{1}{2}$ that $z[i] = x[i]$, and otherwise $z[i] = y[i]$.

With respect to mutation, given a parent $x \in \{0,1\}^n$, the SEMO and the GSEMO differ in how they create offspring $z \in \{0,1\}^n$. Both algorithms first copy x to z and then modify z. The SEMO uses *one-bit mutation*, which flips exactly one bit in z chosen uniformly at random. The GSEMO uses *standard bitwise mutation*, which decides for each position $i \in [n]$ independently whether to flip z_i with probability $\frac{1}{n}$, or whether to leave it unchanged, with the remaining probability.

2.4 THE NSGA-II MULTI-OBJECTIVE OPTIMIZER

The *non-dominated sorting genetic algorithm II* (NSGA-II; Algorithm 2) (Deb et al., 2002) is the most popular MOEA. In contrast to the (G)SEMO, it maintains a population of a fixed, given size $N \in \mathbb{N}_{\geq 1}$. Each iteration, it creates N offspring and selects out of the combination of parent and offspring population N individuals for the following iteration. This selection process follows three

Algorithm 2: The NSGA-II (Deb et al., 2002) with given population size $N \in \mathbb{N}_{\geq 1}$, initial (multiset) population $P_0 \subseteq \{0,1\}^n$ with $|P_0| = N$, and crossover rate $p_c \in [0,1]$, maximizing a given objective function $f\colon \{0,1\}^n \to \mathbb{R}^2$. All set operations operate on multisets.

```
set t ← 0;
for termination criterion not met do
    set Q_t ← ∅;
    for i ∈ [N] do
        with probability p_c do
            choose x, y ∈ P_t uniformly at random with replacement;
            set z to be the result of uniform crossover with parents x and y;
        else
            choose x ∈ P_t uniformly at random;
            set z to be the result of bitwise mutation with parent x;
        set Q_t ← Q_t ∪ {z};
    set R_t = P_t ∪ Q_t;
    let P_{t+1} be a multiset of N individuals from R_t, chosen lexicogaphically by lowest
      non-dominated rank, highest crowding distance, and, last, uniformly at random;
    set t ← t + 1;
return P;
```

steps, detailed below: (1) non-dominated sorting, (2) crowding distance, and (3) uniform random tie-breaker. Selection is performed lexicographically by the three values of these three steps.

Non-dominated sorting partitions a given multiset $R \subseteq \{0,1\}^n$ recursively into *non-dominated levels*. The first level L_1 contains all non-dominated individuals from R, that is, $L_1 = \{x \in R \mid \nexists y \in R\colon y \succ x\}$. Letting $k \in \mathbb{N}_{\geq 1}$ denote the total number of levels, level $i \in [2..k]$ is the multiset of all non-dominated individuals, ignoring levels L_1 to L_{i-1}, that is, $L_i = \{x \in R \setminus \bigcup_{j\in[i-1]} L_j \mid \nexists y \in R \setminus \bigcup_{j\in[i-1]} L_j\colon y \succ x\}$. The index of a level is called the *(non-dominated) rank* of its elements.

Among all ranks, we call the smallest rank $i^* \in [k]$ such that $|\bigcup_{j\in[i^*]} L_j| \geq N$ the *critical* rank. The NSGA-II always chooses all levels L_1 to L_{i^*-1} for the next iteration, but it still needs to select $M \coloneqq N - |\bigcup_{j\in[i^*-1]} L_j|$ individuals from L_{i^*} for the next iteration. This selection is performed with the crowding distance of each individual in L_{i^*}. The crowding distance of each $x \in L_{i^*}$ is the sum of the crowding distance *per objective* of x. The latter is defined for each objective $j \in [2]$ as the normalized distance of x in objective i to its neighbors. Formally, let $(\tilde{y}^{(i,j)})_{i\in[|L_{i^*}|]}$ denote a sorting of L_{i^*} with respect to increasing objective j, breaking ties arbitrarily. The crowding distance per objective j for $\tilde{y}^{(1,j)}$ and $\tilde{y}^{(1,|L_{i^*}|)}$ is positive infinity. For all $i \in [2..|L_{i^*}|-1]$, the crowding distance of $\tilde{y}^{(i,j)}$ is $\big(f_j(\tilde{y}^{(i+1,j)}) - f_j(\tilde{y}^{(i-1,j)})\big)/\big(f_j(\tilde{y}^{(|L_{i^*}|,j)}) - f_j(\tilde{y}^{(1,j)})\big)$. The NSGA-II selects M individuals from L_{i^*} with the largest crowding distance, breaking ties uniformly at random.

3 Discussion of the Previous Theoretical Works

We briefly describe the few previous results and discuss why we are skeptical that they generalize.

As noted in the introduction, there is very little work on how to set the initial population of a MOEA. For empirical works, we refer to Friedrich & Wagner (2015), the references therein, and to the works by Gong et al. (2023; 2025a;b); Benecke & Mostaghim (2025); Benecke (2026). Overall, they suggest a general benefit of initializing MOEAs, besides others, with the single-objective optima.

The first time an initialization with the optima of each objective was proposed for a MOEA is the theoretical work by Qian et al. (2013). Apart from two minor results by Li et al. (2016); Bian et al. (2018), this is also the only work regarding this type of initialization.

Qian et al. (2013) regard a variant of the (G)SEMO. Since all their results are bi-objective problems, we also describe their algorithm only for two objectives. The population is initialized with one optimal solution for each objective. In the main loop of the (G)SEMO, two offspring are created. This

difference to the classic (G)SEMO appears not necessary to us, but to be on the safe side, we describe the algorithm as given by Qian et al. (2013). The algorithm selects two parents x_1 and x_2. From these, two offspring y_1 and y_2 are generated. With probability p_c, which is usually taken as $p_c = \frac{1}{2}$, the two offspring are obtained from a two-offspring one-point crossover, that is, a cut-point $j \in [1 \,..\, n-1]$ is chosen uniformly at random and, for each $i \in \{1, 2\}$, we obtain y_i from concatenating the first j bits from x_i with the last $n - j$ bits from the other parent. With probability $1 - p_c$, we obtain y_i from x_i just via mutation, which is one-bit mutation for the SEMO and bitwise mutation for the GSEMO. We note that the choice of the crossover operator is disputable. For a problem like LOTZ, in which the order of the bits is important, a one-point crossover makes a lot of sense. For problems like ONEMINMAX or COCZ, which are symmetric in the bit-positions, uniform crossover (as used in this work) appears more appropriate. What is a critical difference between the algorithm by Qian et al. (2013) and a typical (G)SEMO is the parent selection. While usually parents are selected randomly from the population, Qian et al. (2013) choose the first parent x_1 randomly among the two optima of the objectives; x_2 is chosen randomly from the rest of the population. Qian et al. (2013) call this a mechanism ensuring high *objective diversity*. The two offspring are then integrated into the population (or not) in the standard (G)SEMO fashion.

The main result by Qian et al. (2013) is that the algorithm described above (when not counting the time to find the optima of the individual objectives) finds the Pareto front of the COCZ and WLPTNO benchmarks (the latter is a generalization of the classic LOTZ benchmark) in an expected time of $O(n \log n)$ function evaluations. This is remarkable in that for the classic (G)SEMO, only the runtime guarantees $O(n^2 \log n)$ and $O(n^3)$ are known (Laumanns et al., 2004; Giel, 2003).

For two reasons, we find it hard to deduce general conclusions from this result, in particular, on the usefulness of the initialization procedure. The first is that the whole setup leads to the very positive optimization behavior that a constant fraction of the offspring are uniformly distributed on the Pareto front (in the sense made precise for ONEMINMAX in the following lemma; analogous proofs would hold for COCZ or WLPTNO). With this property, a simple coupon collector argument over the Pareto-optimal solution values gives the main results (the proof by Qian et al. (2013) is different). The following lemma shows that each offspring of the algorithm variant by Qian et al. (2013) has a probability of order at least $\frac{1}{n}$ for *any* Pareto-optimal objective vector. The reason is that one-point crossover with one of the single-objective optima is particularly useful for the problem(s) studied.

Lemma 1. *Consider one iteration of the algorithm from Qian et al. (2013), as also detailed above, optimizing the* ONEMINMAX *benchmark, abbreviated as f. Let $(k, n-k)$ with $k \in [1 \,..\, n-1]$ be any Pareto-optimal objective vector different from the one of the two initial optima. Let $i \in [1..2]$ and y_i be the i-th offspring generated in this iteration. Then $\Pr[f(y_i) = (k, n-k)] \geq \frac{1}{2} p_c \frac{1}{n-1}$.*

The proofs of the results in this section, which are not central to this work, are in the appendix.

Our second critique on the setup of Qian et al. (2013) is that we find the diversity mechanism very strong. In fact, for the ONEMINMAX benchmark, the diversity mechanism alone, that is, without the initialization with the single-objective optima and without crossover, gives an expected runtime of $O(n \log n)$. Since we do not necessarily have the single-objective optima in the population, we define here the diversity mechanism in the way that the first parent is chosen as the maximizer (in the current population) of a random of the two objectives.

Lemma 2. *Consider the (G)SEMO of Qian et al. (2013), as also detailed above, initialized with a single individual chosen uniformly at random, with $p_c = 0$ (no crossover), and with the diversity mechanism extended to general populations as above. Then the expected time to witness the full Pareto front of the* ONEMINMAX *benchmark is $O(n \log n)$.*

For these two reasons, we find it hard to believe that the main results of Qian et al. (2013) regard a representative setting for a MOEA initialized with the optima of each objective, and that is the main motivation for our work. This does not lessen the general value of the work by Qian et al. (2013). Being the first work proving advantages from crossover on two classic bi-objective benchmarks, it is a cornerstone of the theory of MOEAs, and rightfully its conference version (Qian et al., 2011) has won a best-paper award at the 2011 Genetic and Evolutionary Computation Conference.

Before conducting our runtime analyses, we briefly describe the other known results on MOEAs initialized with single-objective optima. Equally interesting but less related to our work, the initial paper (Qian et al., 2013) also contains a runtime analysis of their algorithm on the bi-objective minimum-spanning-tree problem. Again, a runtime guarantee is proven that is asymptotically faster

than the best known runtime guarantee for the classic GSEMO. Li et al. (2016) show that the SEMO algorithm, with the initialization with the optima of the objectives and without the diversity mechanism and without crossover takes time $\Theta(n^2)$ to find the Pareto front of COCZ. Hence without crossover, there is still a speed-up from the initialization with the single-objective optima, but it is limited to a logarithmic factor. Finally, Bian et al. (2018) regard, as one of the many results shown, a GSEMO with initialization with single-objective optima and a diversity mechanism (different from the one described above). The result is a runtime guarantee for approximating the Pareto front of the newly defined WOMM benchmark. With no other results known for this benchmark, it is not known what the influence of the initialization on this result is.

4 Runtime of SEMO and GSEMO

In this section, we conduct our runtime analysis for the SEMO and GSEMO with uniform crossover (used with constant probability $p_c \in (0,1)$) and an initial population that contains the optima of each objective. We prove the following runtime guarantee for the benchmarks OneMinMax and COCZ.

Theorem 3. *Consider the SEMO or the GSEMO, with uniform crossover, and crossover rate p_c such that p_c and $1-p_c$ are $\Omega(1)$. Assume that the initial population is arbitrary but feasible and contains at least one optimum of each objective. Assume that we optimize the* OneMinMax *or* COCZ *benchmark defined on individuals of length n. Let T be the number of function evaluations performed until the population of the algorithm witnesses the Pareto front. Then T is $O(n^{4/3}\log^{4/3} n)$ both in expectation and with probability at least $1-o(n^{-1})$.*

Given that the following proofs are quite technical, we ease the understanding by regarding only the OneMinMax problem, which is structurally easier, as all solutions are on the Pareto front. However, all our arguments also apply to the COCZ problem since (1) the optima of the objectives are already on the Pareto front, (2) offspring obtained from individuals on the Pareto front lie on the Pareto front again with constant probability, (3) all our proofs remain valid when adding additional individuals to the population (mostly because we always estimate the population size by the largest possible value).

The very rough outer idea of our proof is to alternatingly profit from mutation and crossover. Via mutation, we quickly obtain from a single individual on the Pareto front a longer segment of individuals on the Pareto front (Lemma 5). From individuals from two such longer and distant segments, crossover can construct an individual *in the middle* (Definition 6, as well as Lemmas 15 and 17 in the appendix). Such an individual then creates its own segment, and we repeat our argument. The size ℓ of these segments influences the probability of selecting such parents as well as the time to obtain such segments via mutation. A careful choice of ℓ yields the final runtime guarantee.

To make this rough idea precise, we use the following language. Recall from Section 2.2 that we refer to elements of the Pareto front only by their number of 1-bits. Consequently, we also define intervals on the Pareto front via the number of 1-bits in the corresponding individuals.

Definition 4 (Intervals and Gaps)**.** *Let $P \subseteq \{0,1\}^n$ be a set or multiset, called the population. For an interval $[a..b]$, we call $|[a..b]| = b-a+1$ its* size*. We call an interval $[a..b]$* covered *if for all $i \in [a..b]$, there exists an $x \in P$ with $|x|_1 = i$. We say that $[a..b]$ is a* segment *of the Pareto front if it is an inclusion-maximal covered interval; that is, it is covered and there is no $x \in P$ with $|x|_1 = a-1$ or $|x|_1 = b+1$. Let $\ell \in \mathbb{N}_{\geq 1}$. We say that $[b..c]$ is an ℓ-*gap *if there exist segments $[a\,..\,b-1], [c+1\,..\,d]$ of size at least ℓ and if no segment of size at least ℓ is contained in $[b..c]$. When the value of ℓ is clear from context, we just call $[b..c]$ a* gap*.*

In the remainder, we always assume that the population contains at most n (incomparable) elements, as we are finished once we have $n+1$ incomparable elements, and as any incomparable non-optimal set of individuals has size at most n, both for OneMinMax and COCZ. We first show the following estimate for the time to extend points on the Pareto front into segments of size at least $\ell+1$. This result builds on the fact that with mutation, we obtain an offspring that is a neighbor of the parent and lies (by one) closer to the middle of the Pareto front than the parent with constant probability. To allow for a union bound over several individuals, we also employ a Chernoff bound for sums of geometrically distributed random variables.

Lemma 5. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$. Assume that at some time the population of the algorithm contains elements $x_1, \dots, x_k$. Let $J_1, \dots, J_k$ be the intervals such that, for $i \in [k]$,*

$J_i = [x_i \,..\, x_i + \ell]$ *if* $|x_i|_1 \leq \frac{n}{2}$ *and* $J_i = [x_i - \ell \,..\, x_i]$ *if* $|x_i|_1 > \frac{n}{2}$*. Then with probability at least* $1 - o(n^{-3})$*, all intervals* $J_1, \ldots, J_k$ *are covered after* $4en\ell(1-p_c)^{-1}$ *iterations/function evaluations.*

To analyze the creation of an offspring in a gap, we define the following notion of crossover being centered on a gap, meaning that the number of 1-bits in the offspring lies in the middle half of the gap. Such offspring will later ensure that the size of each gap decreases by a constant factor each time centered crossover is performed.

Definition 6 (Centered Crossover)**.** *Let* $\ell \in \Omega(n^\delta) \cap o(n)$ *for some* $\delta \in (0,1)$ *and let* $[a..b]$ *be an* ℓ*-gap with midpoint* $m = \frac{a+b}{2}$ *and size* $w = b - a + 1$*, and let* $x, y \in \{0,1\}^n$*. Let* Z *be the offspring of uniform crossover of* x *and* y*. We say that* Z *is the result of centered crossover with respect to this gap if* $|Z|_1 \in \left[m - \frac{w}{4}, m + \frac{w}{4}\right]$.

To obtain the best runtime guarantee, we use our general idea of alternating between extending individuals to segments and using crossover to create individuals roughly in the middle of gaps in two different ways. While the gaps are still large *(*1*. Phase)*, we use crossover only from parents chosen from the segments adjacent to a gap. This is sufficient since when the gap is large enough, such a crossover offspring is close to the middle of the gap with high probability. Afterward *(*2*. Phase)*, the gaps are so small that the variance of the position of the crossover offspring reduces the chance that it lies in the desired gap. To compensate for this, we also consider parents that are further away from the gap, profiting from the, at this time, larger population. Unfortunately, the size of the gap at which we need to change from the first to the second phase depends on where this gap lies. Gaps closer to the boundary of the Pareto front suffer less from the sampling variance of crossover but in return allow for fewer suitable pairs of parents. Hence, the distance of the midpoint $\frac{a+b}{2}$ of a gap $[a..b]$ from the boundary elements $\{0, n\}$ is a crucial quantity in the following.

4.1 First Phase

As discussed above, in the first phase we use the alternation of growing segments and creating offspring in the middle of a gap until the gaps are sufficiently small. Precisely, this phase ends when all gaps $[a..b]$ have size at most $\sqrt{m} \ln n$, where m is the distance of the gap's midpoint to the closest boundary of the Pareto front. For the crossover operations, we only consider parents from the two adjacent segments. Since such an offspring is somewhat in the middle of the gap, it (together with the size-ℓ segment obtained from it) reduces the size of the gap by a constant factor. Thus, a logarithmic number of repetitions of this argument ends the first phase. The precise result is the following.

Lemma 7 (First Phase)**.** *Let* $\ell \in \Omega(n^\delta) \cap o(n)$ *for some* $\delta \in (0,1)$ *and let* P *be a population such that the intervals* $[0..\ell]$ *and* $[n - \ell \,..\, n]$ *are covered. Then there exists a constant* $\alpha > 0$ *such that all* ℓ*–gaps* $[a..b]$ *with* $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ *are of size at most* $\max\{4\ell, \sqrt{m} \ln n\}$ *after* $\alpha(\frac{n^2}{\ell^2} \log^2 n + n\ell \log n)$ *iterations/function evaluations with probability* $1 - o(n^{-1})$.

4.2 Second Phase

With the gap sizes now small, we are faced with the difficulty that centered crossover (Definition 6) is increasingly unlikely. We mitigate this by exploiting that now a larger part of the Pareto front is covered. Hence, more pairs of parents are available that can lead to centered crossover. In particular, we use parents not only from the segments directly bordering the gap when their midpoint (the expected place of the crossover offspring) is close enough to the gap's midpoint. Using this way of constructing solutions in gaps and again Lemma 5 to extend such solutions to segments of sufficient size, iteratively a logarithmic number of times, we obtain the following result.

Lemma 8 (Second Phase)**.** *Let* $\ell \in \Omega(n^\delta) \cap o(n)$ *for some* $\delta \in (0,1)$ *and let* P *be a population such that* $[0..\ell]$ *and* $[n - \ell \,..\, n]$ *are covered intervals and that for all* ℓ*–gaps* $[a..b]$ *with* $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$*, the gap has size at most* $\max\{4\ell, \sqrt{m} \ln n\}$*. Then there exists a constant* $\alpha > 0$ *such that it takes at most* $\alpha(\frac{n^2}{\ell^2} \log^2 n + n\ell \log n)$ *iterations/function evaluations with probability* $1 - o(n^{-1})$ *until all* ℓ*–gaps have size at most* 4ℓ*.*

4.3 Final Proof

We now combine the analyses of the first phase (Lemma 7) and the second phase (Lemma 8) into a proof of Theorem 3. This requires first the creation of the initial segments and finally the closing of the remaining small gaps. A restart argument extends the high-probability guarantee to a bound on the expected runtime. We also choose an appropriate value for ℓ to minimize our runtime bound.

Proof of Theorem 3. Define $\ell \coloneqq \lceil \sqrt[3]{n \ln n} \rceil$. With Lemma 5, we can cover the intervals $[0..\ell]$ and $[n-\ell\,..\,n]$ in $O(n\ell) = O(n^{4/3}\log^{1/3} n)$ function evaluations with probability $1-o(n^{-3})$. With this, we have at least one ℓ-gap. Lemma 7 then yields that in time $O(n^2 n^{-2/3}\log^2 n \log^{-2/3} n + nn^{1/3}\log n \log^{1/3} n) = O(n^{4/3}\log^{4/3} n)$ with probability $1-o(n^{-1})$ all ℓ-gaps $[a..b]$ have size at most $\max\{4\ell, \sqrt{m}\log n\}$, where $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Lemma 8 continues from there and gives that in time $O(n^{4/3}\log^{4/3} n)$ with probability $1-o(n^{-1})$ all ℓ-gaps have size at most 4ℓ. To cover those ℓ-gaps, we can apply Lemma 5 again, leading to an additional time of $O(n^{4/3}\log^{1/3} n)$ with probability $1-o(n^{-2})$. This covers the Pareto front. In total, there exists a constant $\alpha > 0$ such that $T < \alpha n^{4/3}\log^{4/3} n$ with probability $1-o(n^{-1})$, proving the high-probability statement.

For the bound on the expectation, we can repeat the above argument, exploiting that our analysis applies to any initial state of the algorithm (that has the single-objective optima in the population). As a shorthand, define $t(n) \coloneqq \alpha n^{4/3}\log^{4/3} n$. If we are not finished in the first $t(n)$ iterations, we can simply start the analysis anew with none of the individuals found previously worsening the runtime. Therefore, for all $i \geq 1$, we have $\Pr[T \geq i\,t(n)] \leq n^{-i}$. Now define $X \coloneqq \lceil \frac{T}{t(n)} \rceil$. Clearly, $T \leq t(n)X$. Since X only takes on positive integers, we can apply a well-known identity for the expected value of non-negative integer random variables, e.g., (Doerr, 2020, Lemma 1.6.1) (Lemma 14 in the appendix), yielding $\mathrm{E}[X] = \sum_{i=1}^{\infty}\Pr[X \geq i] \leq \sum_{i=1}^{\infty}\Pr[T \geq i\,t(n)] \leq \sum_{i=1}^{\infty} n^{-i} \leq \frac{1}{1-\frac{1}{n}} \leq 2$.

Using $\mathrm{E}[T] \leq t(n)\mathrm{E}[X]$, we get $\mathrm{E}[T] \leq 2t(n)$, proving the proposition. □

5 Runtime of NSGA-II

In this section, we analyze the NSGA-II with uniform crossover (see Algorithm 2) on OneMinMax and COCZ for a population size $N \geq 4(n+1)$ that is initialized such that the optima of the single-objective functions are part of the initial population. We obtain the following runtime bound.

Theorem 9. *Regard the NSGA-II with population size $N \geq 4(n+1)$ such that $N \in \Theta(n^{3\delta})$ for some $\delta \in (0,1)$, using uniform crossover with constant probability $p_c \in (0,1)$, and initialized with any population that contains the single-objective optima. Then it takes $O(N^{1/3}\log^{4/3} n)$ iterations, equivalently $O(N^{4/3}\log^{4/3} n)$ function evaluations, both in expectation and with probability $1-o(n^{-1})$, to find the Pareto fronts of the* OneMinMax *or* COCZ *benchmarks.*

For the NSGA-II with random initialization not using crossover, the runtime on the OneMinMax benchmark is known to be $\Theta(Nn\log n)$ (Zheng & Doerr, 2023; Doerr & Qu, 2023a) for reasonable ranges of N. The upper bound here also holds when crossover is used with constant rate less than one. For COCZ, the same upper bound is known (Wietheger & Doerr, 2024). All these upper bounds hold for an arbitrary initialization, hence these bounds also apply to our setting. In this light, the main result of this section shows a better performance from the initialization with the single-objective optima when $N = o(n^3)$. Since all these runtime guarantees deteriorate with increasing value of N, the most interesting case in all results is $N = \Theta(n)$; here we obtain the same asymptotic runtime guarantees and the same speed-up as we did earlier for the (G)SEMO.

For the proof, we note that, as we have $N \geq 4(n+1)$, Lemma 1 by Zheng & Doerr (2023) (which also holds for COCZ, even if this is not stated explicitly) shows that once we have an individual with k 1-bits in the combined parent and offspring population, in all following iterations, the parent population will always contain an individual with k 1-bits. This specifically means that an interval covered (see Definition 4) in an iteration t will be covered in all following iterations. In the following proofs, we make use of this fact without noting it individually each time.

The proof of the theorem is, in spirit, the same as for the (G)SEMO in the previous section, also alternating between using mutation and crossover and gradually covering segments of an appropriately chosen size ℓ and using the same two phases. However, the fixed population size N requires adjustments in some of the proofs, leading to many of the lemmas for this version being slightly altered versions of the lemmas for the (G)SEMO. We show that neither the fixed population size nor the iterations consisting of N offspring creations hinder any of the arguments used for the (G)SEMO.

For a complete overview, we also note a runtime bound of $O(Nn)$ for the NSGA-II with population size $N > 4(n+1)$ when initialized with any population that contains the single-objective optima. This bound is strictly worse than the one in Theorem 9 for $N \in \Theta(n^{3\delta})$, but it holds for larger population sizes. The proof, in the appendix, is similar to the one given in Lemma 5 in that we also build on the constant probability for creating an offspring that lies closer to the middle than the parent.

Lemma 10. *Consider the NSGA-II with population size $N \geq 4(n+1)$ with uniform crossover, and crossover rate p_c such that p_c and $1-p_c$ are $\Omega(1)$. Assume that the initial population is arbitrary, but contains at least one optimum of each objective. Assume we optimize the* OneMinMax *or* COCZ *benchmark defined on individuals of length n. Then, it takes $O(n)$ iterations ($O(Nn)$ function evaluations) both in expectation and with probability at least $1-o(n^{-1})$ until the population of the algorithm witnesses the Pareto front.*

6 Conclusion

This work constitutes the first mathematical analysis of MOEAs initialized with the optima of the individual objectives that do not use non-standard operators such as strong diversity mechanisms. For both the theory-based (G)SEMO algorithm and the in practice dominant NSGA-II, we prove that a population witnessing the Pareto front is found in an expected runtime (number of function evaluations) of $O(n^{4/3}\log^{4/3} n)$ (for $N = \Theta(n)$ for the NSGA-II; Theorems 3 and 9). This is a considerable speed-up from the runtime of order $n^2 \log n$ observed for many standard MOEAs (including (G)SEMO and NSGA-II) without this initialization. Moreover, it is also an excellent runtime for a problem that consists of finding $n+1$ bit strings of length n. Our runtime guarantee is weaker than the $O(n \log n)$ bound proven by Qian et al. (2013) for the (G)SEMO with this initialization and a strong mechanism ensuring objective diversity. We do not see this as a strong limitation of our result. As argued in Section 3, the setting regarded by Qian et al. (2013) is quite specific and leads to very optimistic population dynamics that are rarely seen in other analyses of crossover-based algorithms (Neumann & Theile, 2010; Doerr & Qu, 2023b; Opris, 2026).

What is a clear limitation of our work, shared with most mathematical runtime analyses of MOEAs (e.g., all cited in this work), is that we regard only a few MOEAs on a small number of benchmarks. Given the difficulty of a mathematical runtime analysis and that such an analysis, naturally, has to depend strongly on the particular setting regarded, this seems unavoidable at the moment. However, we are optimistic that our finding that MOEAs can profit from the initialization with the optima of the individual objectives extends to other settings. For example, for the LOTZ benchmark, we are convinced that spreading out from the single-objective optima via mutation and then using crossover between parents from these initial segments will give a better runtime guarantee than the common $O(n^3)$ guarantee. Still, the details will surely be different and need a new proof. This is in particular because for the LOTZ problem, the one-point crossover operator is more sensible than the uniform crossover operator preferred for problems like OneMinMax or COCZ. Moreover, we obtain general valuable insights. For example, our analysis shows how solutions via crossover are generated and can create favorable solutions, such as in the case for OneMinMax and COCZ. If favorable solutions also exist *in the middle* for other problems, standard uniform crossover finds them efficiently. Last, our main proof idea also carries over to settings where the population does not contain optima for each objective but just solutions that are better than uniformly random. The interval surrounded by these solutions is efficiently covered, while the solutions whose objective vectors lie outside this interval still require to be found via mutation. This provides us with a good guideline that an initialization strategy may be particularly useful if the initial solutions can be well obtained by mutation alone.

For all these reasons, we interpret our work as a strong suggestion to add the optima of the individual objectives to the initial population. This may not yield drastic speed-ups in every situation, but it has the potential to do so, and we see little risk that adding these optima to, say, an otherwise large random initial population, will significantly harm the optimization process.

AI USE STATEMENT

This work did not make use of generative AI. It is entirely the result of the work of the author(s), who take full responsibility for the content of this work.

REFERENCES

Tobias Benecke. *Exploring the population dynamics of evolutionary algorithms using gene heritage*. PhD thesis, Otto von Guericke University Magdeburg, 2026.

Tobias Benecke and Sanaz Mostaghim. Tracing genome influence in multi-objective evolutionary algorithms. In *Genetic and Evolutionary Computation Conference Companion, GECCO 2025*, pp. 1807–1815. ACM, 2025. doi: 10.1145/3712255.3734275.

Chao Bian, Chao Qian, and Ke Tang. A general approach to running time analysis of multi-objective evolutionary algorithms. In *International Joint Conference on Artificial Intelligence, IJCAI 2018*, pp. 1405–1411. ijcai.org, 2018.

Chao Bian, Yawen Zhou, Miqing Li, and Chao Qian. Stochastic population update can provably be helpful in multi-objective evolutionary algorithms. In *International Joint Conference on Artificial Intelligence, IJCAI 2023*, pp. 5513–5521. ijcai.org, 2023.

Carlos Artemio Coello Coello, Gary B. Lamont, and David A. van Veldhuizen. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Springer, 2nd edition, 2007.

Duc-Cuong Dang, Andre Opris, and Dirk Sudholt. Crossover can guarantee exponential speed-ups in evolutionary multi-objective optimisation. *Artificial Intelligence*, 330:104098, 2024.

Indraneel Das and John E. Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8: 631–657, 1998.

Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2002.

Anh Viet Do, Aneta Neumann, Frank Neumann, and Andrew M. Sutton. Rigorous runtime analysis of MOEA/D for solving multi-objective minimum weight base problems. In *Advances in Neural Information Processing Systems, NeurIPS 2023*, pp. 36434–36448, 2023.

Benjamin Doerr. Probabilistic tools for the analysis of randomized optimization heuristics. In Benjamin Doerr and Frank Neumann (eds.), *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pp. 1–87. Springer, 2020. Also available at `https://arxiv.org/abs/1801.06733`.

Benjamin Doerr and Zhongdi Qu. From understanding the population dynamics of the NSGA-II to the first proven lower bounds. In *Conference on Artificial Intelligence, AAAI 2023*, pp. 12408–12416. AAAI Press, 2023a.

Benjamin Doerr and Zhongdi Qu. Runtime analysis for the NSGA-II: provable speed-ups from crossover. In *Conference on Artificial Intelligence, AAAI 2023*, pp. 12399–12407. AAAI Press, 2023b.

Benjamin Doerr and Carola Winzen. Ranking-Based Black-Box Complexity. *Algorithmica*, 68 (3):571–609, March 2014. ISSN 1432-0541. doi: 10.1007/s00453-012-9684-9. URL `https://doi.org/10.1007/s00453-012-9684-9`.

Tobias Friedrich and Markus Wagner. Seeding the initial population of multi-objective evolutionary algorithms: A computational study. *Applied Soft Computing*, 33:223–230, 2015.

Oliver Giel. Expected runtimes of a simple multi-objective evolutionary algorithm. In *Congress on Evolutionary Computation, CEC 2003*, pp. 1918–1925. IEEE, 2003.

Oliver Giel and Per Kristian Lehre. On the effect of populations in evolutionary multi-objective optimisation. *Evolutionary Computation*, 18:335–356, 2010.

Cheng Gong, Yang Nan, Lie Meng Pang, Qingfu Zhang, and Hisao Ishibuchi. Effects of including optimal solutions into initial population on evolutionary multiobjective optimization. In *Genetic and Evolutionary Computation Conference, GECCO 2023*, pp. 661–669. ACM, 2023. doi: 10.1145/3583131.3590515.

Cheng Gong, Ping Guo, Lie Meng Pang, Qingfu Zhang, and Hisao Ishibuchi. Population initialization for evolutionary multi-objective optimization: A short review. In *Congress on Evolutionary Computation, CEC 2025*, pp. 1–9. IEEE, 2025a. doi: 10.1109/CEC65147.2025.11043105.

Cheng Gong, Ping Guo, Lie Meng Pang, Qingfu Zhang, and Hisao Ishibuchi. Performance comparison between evolutionary algorithms and linear programming-based relaxation methods for multi-objective knapsack problems. In *Genetic and Evolutionary Computation Conference, GECCO 2025*, pp. 222–230. ACM, 2025b. doi: 10.1145/3712256.3726379.

Pascal Halffmann, Luca E. Schäfer, Kerstin Dächert, Kathrin Klamroth, and Stefan Ruzika. Exact algorithms for multiobjective linear optimization problems with integer variables: A state of the art survey. *Journal of Multi-Criteria Decision Analysis*, 29:341–363, 2022.

Marco Laumanns, Lothar Thiele, Eckart Zitzler, Emo Welzl, and Kalyanmoy Deb. Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem. In *Parallel Problem Solving from Nature, PPSN 2002*, pp. 44–53. Springer, 2002.

Marco Laumanns, Lothar Thiele, and Eckart Zitzler. Running time analysis of multiobjective evolutionary algorithms on pseudo-Boolean functions. *IEEE Transactions on Evolutionary Computation*, 8:170–182, 2004.

Mingfeng Li, Weijie Zheng, and Benjamin Doerr. Scalable speed-ups for the SMS-EMOA from a simple aging strategy. In *International Joint Conference on Artificial Intelligence, IJCAI 2025*, pp. 8885–8893. ijcai.org, 2025.

Yuan-Long Li, Yu-Ren Zhou, Zhi-Hui Zhan, and Jun Zhang. A primary theoretical study on decomposition-based multiobjective evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 20:563–576, 2016.

Frank Neumann and Madeleine Theile. How crossover speeds up evolutionary algorithms for the multi-criteria all-pairs-shortest-path problem. In *Parallel Problem Solving from Nature, PPSN 2010, Part I*, pp. 667–676. Springer, 2010.

Andre Opris. A first runtime analysis of the PAES-25: an enhanced variant of the Pareto Archived Evolution Strategy. In *Foundations of Genetic Algorithms, FOGA 2025*, pp. 202–213. ACM, 2025.

Andre Opris. Many-objective problems where crossover is provably essential. *Artificial Intelligence*, 350:104453, 2026.

Luís Paquete and Thomas Stützle. A two-phase local search for the biobjective traveling salesman problem. In *Evolutionary Multi-Criterion Optimization, EMO 2003*, pp. 479–493. Springer, 2003.

Chao Qian, Yang Yu, and Zhi-Hua Zhou. An analysis on recombination in multi-objective evolutionary optimization. In *Genetic and Evolutionary Computation Conference, GECCO 2011*, pp. 2051–2058. ACM, 2011.

Chao Qian, Yang Yu, and Zhi-Hua Zhou. An analysis on recombination in multi-objective evolutionary optimization. *Artificial Intelligence*, 204:99–119, 2013.

Shengjie Ren, Chao Bian, Miqing Li, and Chao Qian. A first running time analysis of the Strength Pareto Evolutionary Algorithm 2 (SPEA2). In *Parallel Problem Solving from Nature, PPSN 2024, Part III*, pp. 295–312. Springer, 2024.

Shengjie Ren, Zimin Liang, Miqing Li, and Chao Qian. Not just for archiving: provable benefits of reusing the archive in evolutionary multi-objective optimization. In *Conference on Artificial Intelligence, AAAI 2026*, pp. 37152–37160. AAAI Press, 2026.

Simon Wietheger and Benjamin Doerr. A mathematical runtime analysis of the non-dominated sorting genetic algorithm III (NSGA-III). In *International Joint Conference on Artificial Intelligence, IJCAI 2023*, pp. 5657–5665. ijcai.org, 2023.

Simon Wietheger and Benjamin Doerr. Near-tight runtime guarantees for many-objective evolutionary algorithms. In *Parallel Problem Solving from Nature, PPSN 2024, Part IV*, pp. 153–168. Springer, 2024.

Weijie Zheng and Benjamin Doerr. Mathematical runtime analysis for the non-dominated sorting genetic algorithm II (NSGA-II). *Artificial Intelligence*, 325:104016, 2023.

Weijie Zheng, Yufei Liu, and Benjamin Doerr. A first mathematical runtime analysis of the non-dominated sorting genetic algorithm II (NSGA-II). In *Conference on Artificial Intelligence, AAAI 2022*, pp. 10408–10416. AAAI Press, 2022.

A Appendix for "Initializing Multi-Objective Evolutionary Algorithms with the Optima of Each Objective"

You may include other additional sections here.

In this appendix, we provide supplementary information to be read by the reviewers should they wish so. In particular, we state all mathematical proofs that were omitted in the body of the submission for reasons of space. For the sake of convenience, we repeat the statements that we prove with the same number as in the main paper.

A.1 Preliminaries—Probability Theory

We use the following well-known results in our proofs. The following two theorems show strong concentration of sums of independent random variables.

Theorem 11 (Additive Chernoff Bound from Doerr (2020), Theorem 1.10.7)**.** *Let $X_1, \dots, X_n$ be independent random variables taking values in $[0,1]$. Let $X \coloneqq \sum_{i=1}^{n} X_i$. Then for all $\lambda \geq 0$,*

$$\Pr[X \geq \mathrm{E}[X] + \lambda] \leq \exp\left(-\frac{2\lambda^2}{n}\right).$$

Theorem 12 (Chernoff Bound for Geometric Random Variables from Doerr (2020), Theorem 1.10.32)**.** *Let $X_1, \dots, X_n$ be independent geometric random variables with common success probability $p > 0$. Let $X \coloneqq \sum_{i=1}^{n} X_i$ and $\mu \coloneqq \mathrm{E}[X] = \frac{n}{p}$. For all $\delta \geq 0$,*

$$\Pr[X \geq (1+\delta)\mu] \leq \exp\left(-\frac{1}{4}\min\{\delta^2, \delta\}(n-1)\right).$$

The following lemma states that the probability to be within a certain range of the standard deviation of a standard binomial random variable is lower bounded by a uniform distribution over this interval, up to constant factors.

Lemma 13 (Paraphrased from Doerr & Winzen (2014), Lemma 8)**.** *Let $n \in \mathbb{N}, \kappa \geq 2$ and $\ell \in [\frac{n}{2} \pm \kappa\sqrt{n}]$, and let $X \sim \mathrm{Bin}(n, \frac{1}{2})$. For large enough n we have $\Pr[X = \ell] \geq \beta n^{-1/2}$ for some constant $\beta > 0$.*

The last lemma is a well-known simple reformulation of the expected value for non-negative integer random variables.

Lemma 14 ((Doerr, 2020), Lemma 1.6.1)**.** *Let X be a random variable taking values in the non-negative integers. Then $\mathrm{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]$.*

A.2 Material for Section 3

Lemma 1. *Consider one iteration of the algorithm from Qian et al. (2013), as also detailed above, optimizing the* OneMinMax *benchmark, abbreviated as f. Let $(k, n-k)$ with $k \in [1\,..\,n-1]$ be any Pareto-optimal objective vector different from the one of the two initial optima. Let $i \in [1..2]$ and y_i be the i-th offspring generated in this iteration. Then $\Pr[f(y_i) = (k, n-k)] \geq \frac{1}{2} p_c \frac{1}{n-1}$.*

Proof. We prove the result for $i = 1$ and note that analogous arguments solve the case $i = 2$. Let x_2 be the second parent chosen. Assume first that $f_1(x_2) < k$, so the number of zeros in x_2, is smaller than k. Let $j \in [1\,..\,n-1]$ be such that the first j bits of x_2 contain $k - f_1(x_2)$ ones (such a position exists since x_2 contains $n - f_1(x_2) > k - f_1(x_2)$ ones). Now with probability p_c, our iteration performs crossover, with probability $1/2$, x_1 is the all-zero string (the optimum of the first objective), and with probability $\frac{1}{n-1}$ the cut-point is chosen equal to j. In this case, y_1 has exactly k zeros (that is $f(y) = (k, n-k)$) since the $k - f_1(x_2)$ ones in the first j bits of x_2 have been overwritten by zeros. If $f_1(x_2) > k$, an analogous argument with the all-ones string as x_1 shows the claim. When $f_1(x_2) = k$, then we do the same argument with x_1 being the string having all bits equal to the first bit of x_2 and the cut-point $j = 1$, resulting in an offspring y_1 that is equal to x_2. □

Lemma 2. *Consider the (G)SEMO of Qian et al. (2013), as also detailed above, initialized with a single individual chosen uniformly at random, with $p_c = 0$ (no crossover), and with the diversity mechanism extended to general populations as above. Then the expected time to witness the full Pareto front of the* ONEMINMAX *benchmark is $O(n \log n)$.*

Proof. Let f denote ONEMINMAX. We start with the easier case of the SEMO, which already contains the main working principles of this algorithm. Since the SEMO performs one-bit mutation, the population always forms an interval of the Pareto front. If the current maximizer of f_1 in the population has k zeros, then with probability $1/2$ the first parent x_1 is this maximizer, with probability $\frac{n-k}{n}$ mutation flips a one to a zero, and the maximum f_1 value in the population is increased. Hence the expected waiting time for this event is $2\frac{n}{n-k}$. An analogous waiting time can be shown for increasing the maximum f_2 value. Summing over these waiting times (that is, pessimistically summing up these terms for all $k \in [1..n]$) gives an expected time of $O(n \log n)$ for witnessing the Pareto front.

For the GSEMO, the times to increase the maximum objective values are similar, higher by at most a factor of $(1-1/n)^{-(n-1)} \leq e \in O(1)$ for conditioning on the event that exactly one bit flips. Hence again after an expected number of $O(n \log n)$ iterations, the two extremal points of the Pareto front are found. Using a Chernoff bound for sums of geometrically distributed random variables with harmonic expectations (here the waiting times discussed above), e.g., Theorem 1.10.35 by Doerr (2020), we also have this result with probability at least $1 - n^{-3}$.

The crucial difference from the SEMO is that we may increase an objective value by more than one, leaving a gap in the Pareto front. These gaps can be at most as wide as the number of bits flipped in the corresponding iteration. Since the number of bits flipped follows a binomial distribution with parameters n and $1/n$, in particular, having an expectation of one, a Chernoff bound for large deviations over the mean such as Corollary 1.10.4. by Doerr (2020), shows that the probability that a single mutation flips more than $5 \log_2(n)$ bits is at most n^{-5}. Using a union bound over the $O(n \log n)$ iterations, we obtain that with probability at least $1 - O(n^{-4} \log n)$, all gaps have size at most $5 \log_2 n \in O(\log n)$. The probability to reduce the length of such a gap by simply choosing the outer end point as parent and mutating it to the adjacent objective value in the gap is at most $\frac{1}{n+1}\frac{1}{2}$ analogous to the situation discussed in Lemma 5. Hence again analoguos to the proof of this lemma, we can fill all these gaps of at most logarithmic length in time $O(n \log n)$ with probability $1 - o(n^{-3})$. In total, with probability $1 - O(n^{-3})$, we find the Pareto front in time $O(n \log n)$.

In the rare failure case of this argument (occurring with probability $O(n^{-3})$), we can adapt the classic $O(n^2 \log n)$ runtime proof for the GSEMO on ONEMINMAX to our situation. For this we note that also in the presence of the diversity mechanism, in each iteration each member of the population has a probability of at least $\frac{1}{n+1}$ of being chosen at least once as parent for mutation. Hence also in this rare failure case, an additional expected number of $O(n^2 \log n)$ iterations finds the full Pareto front. Combining the regular and the failure case with the law of total expectation, we obtain an expected time of $O(n \log n)$ function evaluations (and iterations) for the GSMEO to find the Pareto front. □

A.3 MATERIAL FOR SECTION 4

A.3.1 PROOF OF THE SEGMENT-GROWING ARGUMENT

Lemma 5. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$. Assume that at some time the population of the algorithm contains elements $x_1, \dots, x_k$. Let $J_1, \dots, J_k$ be the intervals such that, for $i \in [k]$, $J_i = [x_i \,..\, x_i + \ell]$ if $|x_i|_1 \leq \frac{n}{2}$ and $J_i = [x_i - \ell \,..\, x_i]$ if $|x_i|_1 > \frac{n}{2}$. Then with probability at least $1 - o(n^{-3})$, all intervals $J_1, \dots, J_k$ are covered after $4en\ell(1-p_c)^{-1}$ iterations/function evaluations.*

Proof. Fix an interval J_i and let x be an individual with $|x|_1 \in J_i$. By symmetry, we may assume that $|x|_1 \leq \frac{n}{2}$. Then $|x|_0 \geq \frac{n}{2} - \ell$. For sufficiently large n, $\ell < \frac{n}{6}$. It follows that $|x|_0 \geq \frac{n}{3}$, so the probability of flipping such a bit is at least $\frac{1}{3}$, conditioned on choosing mutation and then flipping exactly one bit. For GSEMO, this happens with probability at least $\frac{1}{e}(1-p_c)$, whereas for the SEMO, it happens with probability $1 - p_c \geq \frac{1}{e}(1-p_c)$. Pessimistically assuming that we cover the interval sequentially, the time to cover it is stochastically dominated by a sum of ℓ geometric random variables with success probabilities $\frac{1}{3e}(1-p_c)n^{-1}$, since the population size is at most n.

Denote these variables by $G_1, \dots, G_\ell$ and their sum by $X_i \coloneqq \sum_{j=1}^{\ell} G_j$. Then we have $\mathrm{E}[X_i] = 3en\ell(1-p_c)^{-1}$. We may consider the intervals independently, as any influence these processes have on each other will not worsen the runtime.

By Theorem 12 we have

$$\Pr\left[X_i \geq \left(1+\frac{1}{3}\right) \cdot 3en\ell(1-p_c)^{-1}\right] \leq \exp\left(-\frac{1}{36}(\ell-1)\right) \leq e\exp\left(-\frac{1}{36}\ell\right) \in o(n^{-4}).$$

The event that any of these intervals are not covered within $4en\ell(1-p_c)^{-1}$ steps is the union of any single one not being covered in that time. Therefore, by the union bound, it has probability at most $k \cdot o(n^{-4}) \subseteq n \cdot o(n^{-4}) = o(n^{-3})$. □

A.3.2 Analysis of Phase 1

For the analysis of Phase 1, we first need an argument showing that sufficiently large gaps are hit by centered crossover of parents from the neighboring segments with sufficiently high probability.

Lemma 15. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let $[a..b]$ be an ℓ-gap with size $w = b-a+1$ and $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Assume that $w \geq \max\{4\ell, \sqrt{m}\ln n\}$.*

Let x, y be individuals with $|x|_1 \in [a-\ell \,..\, a-1]$ and $|y|_1 \in [b+1 \,..\, b+\ell]$, and let Z be the offspring of uniform crossover of x and y. Then with probability at least $1 - 2 \cdot n^{-\frac{1}{128}\ln n}$, the individual Z is the result of centered crossover with respect to the gap.

Proof. By symmetry, we may assume that $\frac{a+b}{2} \leq \frac{n}{2}$. Let $m' = \frac{|x|_1+|y|_1}{2}$ be the midpoint of the number of 1-bits of x and y. As we choose x and y with a number of 1-bits from $[a-\ell \,..\, a-1]$ and $[b+1 \,..\, b+\ell]$ respectively, and since $\ell \leq \frac{w}{4}$, we have

$$m' \in \left[m - \frac{\ell}{2}, m + \frac{\ell}{2}\right] \subseteq \left[m - \frac{w}{8}, m + \frac{w}{8}\right].$$

The probabiliy of Z not being the result of centered crossover therefore is

$$\Pr\left[|Z|_1 \notin \left[m - \frac{w}{4}, m + \frac{w}{4}\right]\right] \leq \Pr\left[|Z|_1 \notin \left[m' - \frac{w}{8}, m' + \frac{w}{8}\right]\right].$$

Let s be the number of positions in which both x and y have a 1-bit. Let $W = |Z|_1 - s$. Then $W \sim \mathrm{Bin}(H(x,y), \frac{1}{2})$. We have $\mathrm{E}[W] = \frac{1}{2}H(x,y) = \frac{1}{2}(|x|_1 + |y|_1 - 2s) = m' - s$.

Note that

$$H(x,y) \leq |x|_1 + |y|_1 \leq a + b + \ell \leq a + b + a \leq 2(a+b) = 4m.$$

Using Theorem 11, we compute

$$\begin{aligned}\Pr\left[|Z|_1 \geq m' + \frac{w}{8}\right] = \Pr\left[W \geq \mathrm{E}[W] + \frac{w}{8}\right] &\leq \exp\left(-\frac{2\left(\frac{w}{8}\right)^2}{H(x,y)}\right) = \exp\left(-\frac{1}{32} \cdot \frac{(w)^2}{H(x,y)}\right) \\ &\leq \exp\left(-\frac{1}{32} \cdot \frac{m\ln^2 n}{4m}\right) = \exp\left(-\frac{1}{128} \cdot \ln^2 n\right) = n^{-\frac{1}{128}\ln n}.\end{aligned}$$

Due to the symmetry of a binomial distribution with success probability $\frac{1}{2}$, we also have

$$\Pr\left[|Z|_1 \leq m' - \frac{w}{8}\right] = \Pr\left[W \leq \mathrm{E}[W] - \frac{w}{8}\right] = \Pr\left[W \geq \mathrm{E}[W] + \frac{w}{8}\right] \leq n^{-\frac{1}{128}\ln n}.$$

Putting it all together, we obtain a probability of Z not being the result of centered crossover of at most $2 \cdot n^{-\frac{1}{128}\ln n}$. □

We now use suitable crossover from this result as well as the segment-growing argument of Lemma 5 to analyze the time it takes to decrease certain gap sizes by a constant factor.

Lemma 16. *Let $w \in \mathbb{N}$ and $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$. Let P be a population such that all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, whose size is at least $\max\{4\ell, \sqrt{m}\ln n\}$, have size at most w. Then, with probability at least $1-o(n^{-2})$, after at most $4p_c^{-1}\frac{n^2}{\ell^2}\ln n + 4en\ell(1-p_c)^{-1}$ iterations (and function evaluations), all such gaps have size at most $\frac{3}{4}w$.*

Proof. Let $[a..b]$ be an ℓ-gap of size at most w and $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Due to symmetry, we may assume that $m = \frac{a+b}{2}$. Since we are not interested in gaps of size less than $\max\{4\ell, \sqrt{m}\ln n\}$, assume that $w > \max\{4\ell, \sqrt{m}\ln n\}$.

We are now interested in the time it takes to split $[a..b]$ into at least two gaps, each of size at most $\frac{3}{4}w$.

First, note that performing centered crossover on a gap $[a..b]$, yielding a new individual x with $c \coloneqq |x|_1 \in [m - \frac{w}{4}, m + \frac{w}{4}]$, we obtain two intervals $[a\,..\,c-1]$ and $[c+1\,..\,b]$, neither of which containing a segment of size ℓ and each of size at most $\frac{3}{4}w$. If this centered crossover is followed by either covering the interval $[c\,..\,c+\ell]$ or the interval $[c-\ell\,..\,c]$, we obtain at least two gaps, each of size at most $\frac{3}{4}w$.

Regarding the time for centered crossover, let $[a_1..b_1], \dots, [a_k..b_k]$ be the gaps at a given time. We know each gap to be bordered by segments of size at least ℓ by definition. Therefore, for each gap, the probability of choosing one parent each from the ℓ adjacent elements to either side of the gap is at least $p_c\frac{\ell^2}{n^2}$ per offspring generation. Thus, the probability that we need to wait more than $4p_c^{-1}\frac{n^2}{\ell^2}\ln n$ objective function evaluations until a given gap has produced offspring by crossover is

$$\left(1 - \frac{p_c\ell^2}{n^2}\right)^{4p_c^{-1}\frac{n^2}{\ell^2}\ln n} \leq \exp(-4\ln n) \in o(n^{-3})$$

According to Lemma 15, the probability that that crossover is not centered is at most $2 \cdot n^{-\frac{1}{128}\ln n}$. Therefore, the probability, that no centered crossover has been performed on the gap in the given time is in $o(n^{-3})$. A union bound over all $k \leq n$ gaps yields a probability in $o(n^{-2})$ of this step taking more than $4p_c^{-1}\frac{n^2}{\ell^2}\ln n$ objective function evaluations.

Lemma 5 bounds the time for covering intervals of size ℓ to be on the order of $4en\ell(1-p_c)^{-1}$ with probability $1 - o(n^{-3})$. Taken together, reducing the size of all gaps to at most $\frac{3}{4}w$ takes at most $4p_c^{-1}\frac{n^2}{\ell^2}\ln n + 4en\ell(1-p_c)^{-1}$ objective function evaluations with probability $1 - o(n^{-2})$. □

We now derive the runtime bound for our first phase, setting a value for ℓ and bounding the number of repeated applications of Lemma 16 until all gaps are sufficiently small.

Lemma 7 (First Phase)**.** *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let P be a population such that the intervals $[0..\ell]$ and $[n-\ell\,..\,n]$ are covered. Then there exists a constant $\alpha > 0$ such that all ℓ–gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ are of size at most $\max\{4\ell, \sqrt{m}\ln n\}$ after $\alpha(\frac{n^2}{\ell^2}\log^2 n + n\ell\log n)$ iterations/function evaluations with probability $1 - o(n^{-1})$.*

Proof. We can apply Lemma 16 to reduce the maximum gap size by a factor of $\frac{3}{4}$. Each of those reductions takes at most $4p_c^{-1}\frac{n^2}{\ell^2}\ln n + 4en\ell(1-p_c)^{-1}$ iterations with probability $1 - o(n^{-2})$. This is $O(\frac{n^2}{\ell^2}\log n + n\ell)$ objective function evaluations with probability $1 - o(n^{-2})$.

After applying Lemma 16 on the order of $\ln n$ times, all gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ have size at most $\max\{4\ell, \sqrt{m}\ln n\}$. A union bound over the failure probabilities gives us a probability of $1 - o(n^{-1})$ of finishing within $O(\frac{n^2}{\ell^2}\log^2 n + n\ell\log n)$ objective function evaluations. □

A.4 Phase 2

To prove the main result of this phase, we first regard the probability of a single crossover operation from appropriately chosen parent individuals being centered. We observe that due to the higher variance of the position of the crossover offspring this probability is smaller towards the center of the Pareto front.

Lemma 17. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$ and let $[a..b]$ be an ℓ-gap of size at least 4ℓ. Let $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ and let $[a..b]$ be of size in $[(\frac{3}{4})^{\frac{1}{2}} q\sqrt{m}\ln n, q\sqrt{m}\ln n]$ for some $q \leq 1$. Further, let x, y be individuals in the population such that there exists an $r \leq \frac{1}{4}q\sqrt{m}\ln n$ such that $\frac{|x|_1 + r + |y|_1}{2} = \frac{a+b}{2}$.*

Let Z be the offspring of uniform crossover of x and y. Let A be the event that Z is the result of centered crossover on $[a..b]$. Then, for sufficiently large n, there exists a constant $\beta > 0$ such that $\Pr[A] \geq \beta(2m)^{-\frac{1}{2}}\sqrt{\ell}$.

Proof. By symmetry, we may assume that $m = \frac{a+b}{2}$, meaning the distance to the closest edge of the Pareto front corresponds to the number of 1-bits of the midpoint. Further assume w.l.o.g. that $|y|_1 > |x|_1$.

Let s be the number of 1-bits x and y agree on. Then $|Z|_1 \sim s + \mathrm{Bin}(H(x,y), \frac{1}{2})$.

Let $d = \min\{|m - |x|_1|, |m - |y|_1|\}$ and w.l.o.g., assume $d = |m - |y|_1|$, so y is closer to the midpoint of $[a..b]$, which is m, than x is. We know that x has $m - d - r$ 1-bits and y has $m + d$ 1-bits, and the number of 1-bits they disagree in corresponds to the Hamming distance, so $H(x,y) = m - d - r - s + m + d - s = 2m - 2s - r$.

Let $W = |Z|_1 - s \sim \mathrm{Bin}(H(x,y), \frac{1}{2})$. We have $\mathrm{E}[W] = \frac{1}{2}H(x,y) = \frac{1}{2}(2m - 2s - r) = m - s - \frac{r}{2}$.

We first show that W taking on a value that lies in $[\mathrm{E}[W] \,..\, \mathrm{E}[W] + \sqrt{\ell}]$ implies Z being the result of centered crossover. In the next step, we then use this fact to give a bound on the probability for W taking on a value in that interval, which is therefore likewise a lower bound on the probability of Z being the result of centered crossover.

First, recall that $|Z|_1 = s + W$, so for all values of W that lie in $[\mathrm{E}[W] \,..\, \mathrm{E}[W] + \sqrt{\ell}]$, we have $|Z|_1 \in [\mathrm{E}[W] + s \,..\, \mathrm{E}[W] + s + \sqrt{\ell}]$, so $|Z|_1 \in [m - \frac{r}{2} \,..\, m + \sqrt{\ell} - \frac{r}{2}]$. We have $0 \leq r \leq \frac{1}{4}q\sqrt{m}\ln n$, so

$$|Z|_1 \in \left[m - \frac{\frac{1}{4}q\sqrt{m}\ln n}{2} \,..\, m + \sqrt{\ell} - \frac{0}{2}\right] = \left[m - \frac{\frac{1}{2}q\sqrt{m}\ln n}{4} \,..\, m + \sqrt{\ell}\right].$$

Note that $\frac{1}{2}q\sqrt{m}\ln n < (\frac{3}{4})^{\frac{1}{2}}q\sqrt{m}\ln n$, which is the minimum size for $[a..b]$. With this, $|Z|_1$ lies within the bounds for centered crossover for $[a..b]$.

Let $w \in [\mathrm{E}[W] \,..\, \mathrm{E}[W] + \sqrt{\ell}]$. Note that $H(x,y) \geq |y|_1 - |x|_1 \geq 4\ell$, and recall $W \sim \mathrm{Bin}(H(x,y), \frac{1}{2})$, so $\mathrm{E}[W] = \frac{1}{2}H(x,y) \geq 2\ell > \ell$. This means that $\sqrt{\mathrm{E}[W]} > \sqrt{\ell}$, so $[\mathrm{E}[W] \,..\, \mathrm{E}[W] + \sqrt{\ell}] \subseteq [\mathrm{E}[W] - \sqrt{\mathrm{E}[W]} \,..\, \mathrm{E}[W] + \sqrt{\mathrm{E}[W]}]$. We can therefore apply Lemma 13, so we know that $\Pr[W = w] \geq \beta H(x,y)^{-\frac{1}{2}}$ for a constant $\beta > 0$. Summed up over all $\sqrt{\ell}$ possible values w, the probability that $W \in [\mathrm{E}[W] \,..\, \mathrm{E}[W] + \sqrt{\ell}]$ is then at least $\beta H(x,y)^{-\frac{1}{2}}\sqrt{\ell} \geq \beta(2m)^{-\frac{1}{2}}\sqrt{\ell}$. This is likewise a bound on the probability of Z being the result of centered crossover on $[a..b]$. □

We now use this result to prove a runtime bound for performing centered crossover on a given gap $[a..b]$ with midpoint m. Here, we use two main aspects: First, we need to increase the probability of any given crossover operation being centered. For this, we can perform crossover $\sqrt{m}$ times, proving that this leads to a low failure probability. However, it also adds a factor of $\sqrt{m}$ to the runtime. This is canceled out by the parent selection, for which we make use of the fact that the smaller interval size implies that much of the Pareto front is already covered. This is especially beneficial for gaps towards the center of the Pareto front, as they have many possible parent individuals on both sides. This gives us by a factor of $\sqrt{m}$ more parent individuals to choose from for one parent than in the proof for phase one, where we considered only adjacent segments.

Lemma 18. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$ and let $[a..b]$ be an ℓ-gap. Further, let $q \leq 1$ and, for all ℓ-gaps $[e..f]$ with size w_{ef} and $m_{ef} = \min\{\frac{e+f}{2}, n - \frac{e+f}{2}\}$, assume $4\ell \leq w_{ef} \leq q\sqrt{m_{ef}}\ln n$. Then, with probability in $1 - o(n^{-3})$ it takes at most $160\sqrt{3} \cdot q \cdot \frac{n^2}{p_c \ell^2}\ln^2 n$ iterations (objective function evaluations) for a constant probability p_c to perform centered crossover on $[a..b]$.*

Proof. Let $m = \min\{\frac{a-b}{2}, n - \frac{a-b}{2}\}$. By symmetry, we may assume that $m = \frac{a+b}{2}$, meaning the distance to the closest edge of the Pareto front corresponds to the number of 1-bits of the midpoint.

We first want to choose individuals x and y to perform crossover on. We start by regarding the number of choices we have for each.

First, we want to choose an individual y from the interval $[b + \frac{15}{16}m \,..\, b + m]$. We distinguish two cases. For the first one, assume an ℓ-gap with its midpoint in $[b + \frac{15m}{16} .. b + m]$ exists. Take any such ℓ-gap $[e..f]$. Then, the size of the gap $[e..f]$ is at most

$$q\sqrt{\frac{e+f}{2}}\ln n \leq q\sqrt{\frac{2(b+m)}{2}}\ln n = q\sqrt{b+m}\ln n \leq q\sqrt{3m}\ln n,$$

since $b \leq 2m$. We know that segments $[e' \,..\, e-1], [f+1 \,..\, f']$ of size at least ℓ must exist. Therefore, the existence of any ℓ-gap implies the existence of one segment of size at least ℓ. For a lower bound on the number of covered elements, assume that each of those segments has size exactly ℓ. Regarded over the $\frac{m}{16}$ elements in $[b + \frac{15m}{16} \,..\, b + m]$, this gives us at least $\ell \cdot \frac{\frac{m}{16}}{q\sqrt{3m}\ln n + \ell}$ individuals. With $\sqrt{3} \cdot q \ln n > 4\ell$, this is greater than $\ell \cdot \frac{\frac{m}{16}}{2q\sqrt{3m}\ln n} = \ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q \ln n}$. With this, we get at least $\ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q \ln n}$ covered elements in $[b + \frac{15m}{16} \,..\, b + m]$ to choose y from.

For the second case, assume no ℓ-gap $[e..f]$ with $\frac{e+f}{2} \in [b + \frac{15m}{16} \,..\, b + m]$ exists. Let $[g..h]$ be any gap such that $[b + \frac{15m}{16} \,..\, b + m] \cap [g..h] \neq \emptyset$ and let $m_{gh} = \min(\frac{g+h}{2}, n - \frac{g+h}{2}) \leq \frac{g+h}{2}$. Then, $g \leq b + \frac{15m}{16}$. Let w_{gh} be the size of $[g..h]$. Then, we know that $m_{gh} \leq \frac{g+h}{2} = \frac{2g+w_{gh}}{2} \leq b + \frac{15m}{16} + \frac{w_{gh}}{2}$. With $w_{gh} \leq q\sqrt{m_{gh}}\ln n$, we have $w_{gh} \leq m_{gh}$ for sufficiently large n and with $m_{gh} > \ell$. With this and $b \leq 2m$, we have $m_{gh} \leq b + \frac{15m}{16} + \frac{w_{gh}}{2} \leq 2m + \frac{15m}{16} + \frac{m_{gh}}{2}$, so $m_{gh} \leq 2(2m + \frac{15m}{16})$. Accordingly, we get $w_{gh} \leq q\sqrt{2(2m + \frac{15m}{16})}\ln n$. For sufficiently large n, this is less than $\frac{1}{4}$ of $[b + \frac{15m}{16} \,..\, b + m]$. Therefore, with at most two such gaps (one from either side of the interval), at most half of $[b + \frac{15m}{16} \,..\, b + m]$ is part of an ℓ-gap for sufficiently large n, giving us at least $\frac{1}{2}\frac{m}{16}$ elements to choose from for y. Combined with the other case and with $m > \ell$, we get at least $\ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q \ln n}$ elements to choose from for y.

Let $y \in P$ with $|y|_1 \in [b + \frac{15m}{16} \,..\, b + m]$. Let $d \in \mathbb{N}$ be the distance of y to m, meaning $|y|_1 = m + d$. For choosing x, we want at least ℓ possible elements with the number of 1-bits as close to $m - d$ as possible. However, it is possible for $m - d$ to be part of a gap. Call that gap $[g..h] \subseteq [0 \,..\, a - \frac{15m}{16}]$. Its size is then at most

$$q\ln(n)\sqrt{\frac{g+h}{2}} \leq q\ln(n)\sqrt{\frac{2(a - \frac{15m}{16})}{2}} \leq q\ln(n)\sqrt{\frac{m}{16}} = \frac{1}{4}q\ln(n)\sqrt{m}.$$

Therefore, for an $r \leq \frac{1}{4}q\ln(n)\sqrt{m}$, there exists a segment $[z \,..\, m - d - r]$ of size at least ℓ. We select x from the elements in $[m - d - r - \ell \,..\, m - d - r]$ of this segment.

With this, we have at least $\ell \cdot \frac{\sqrt{m}}{2\sqrt{3}\cdot q \ln n}$ choices for selecting x and at least ℓ choices for selecting y. With a population size of at most n, this gives us a probability of at least $\frac{\ell^2\sqrt{m}}{32\sqrt{3}\cdot q n^2 \ln n}$ of selecting x and y as described. Together with a crossover rate p_c, we get a total probability of performing crossover on these elements of $\frac{p_c\ell^2\sqrt{m}}{32\sqrt{3}\cdot q n^2 \ln n}$.

With this, the probability that any such pair has not been selected for crossover in the first $160\sqrt{3} \cdot q \cdot \frac{n^2}{p_c\ell^2\sqrt{m}} \ln^2 n$ steps is

$$\left(1 - \frac{p_c\ell^2\sqrt{m}}{32\sqrt{3} \cdot q\ln(n)n^2}\right)^{160\sqrt{3}\cdot q \cdot \frac{n^2}{p_c\ell^2\sqrt{m}}\ln^2 n} \leq \exp(-5\ln n) \in o\big(n^{-4}\big).$$

By Lemma 17, we have a probability of at least $\beta(2m)^{-\frac{1}{2}}\sqrt{\ell}$ of the performed crossover operation being centered on $[a..b]$. Performing $\sqrt{m}$ crossover operations, the probability of none of them being

centered is

$$\left(1 - \frac{\beta\sqrt{\ell}}{\sqrt{2m}}\right)^{\sqrt{m}} \leq \exp\left(-\frac{\beta}{\sqrt{2}}\sqrt{\ell}\right).$$

This means we have a probability of at most $\exp(-\beta(\frac{\ell}{2})^{\frac{1}{2}})$ of not performing centered crossover in $\sqrt{m}$ crossover operations, conditioned on having selected parents x and y as described for each crossover. A union bound over $\sqrt{m} \leq n$ crossover operations gives a probability of the parent selection for all crossover operations combined taking longer than $160\sqrt{3} \cdot \sqrt{m} \cdot q \cdot \frac{n^2}{p_c \ell^2 \sqrt{m}} \ln^2 n = 160\sqrt{3} \cdot q \cdot \frac{n^2}{p_c \ell^2} \ln^2 n$ objective function evaluations in $o(\sqrt{m}n^{-4}) \subseteq o(n^{-3})$. With $\exp(-\beta(\frac{\ell}{2})^{\frac{1}{2}}) \in o(n^{-3})$, the probability of not performing centered crossover on $[a..b]$ in at most $160\sqrt{3} \cdot q \cdot \frac{n^2}{p_c \ell^2} \ln^2 n$ objective function evaluations is in $o(n^{-3})$. □

We now combine this result with Lemma 5 to give a runtime bound for decreasing the maximum gap size by a constant factor.

Lemma 19. *Let $i \in \mathbb{N}$ and let $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$. Let P be a population such that for all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, the gap has size at most $\max\{4\ell, (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n\}$. Then, with probability in $1 - o(n^{-1})$ it takes at most $\frac{160}{p_c}\sqrt{3}(\frac{3}{4})^{\frac{i}{2}} \cdot \frac{n^2}{\ell^2}\ln^2 n + 4en\ell(1-p_c)^{-1}$ iterations (objective function evaluations) for a constant probability p_c until all ℓ-gaps have size at most $\max\{4\ell, (\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m}\ln n\}$.*

Proof. Let $[a..b]$ be an ℓ-gap with size w and $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Due to symmetry, we may assume that $m = \frac{a+b}{2}$. Since we are only interested in ℓ-gaps of size greater than 4ℓ, assume that $4\ell < w \leq (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n$. Likewise, assume that $w \geq (\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m}\ln n$, as otherwise, $[a..b]$ already fulfills the condition.

We are now interested in the time it takes to split $[a..b]$ into at least two gaps, each of size at most $(\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m'}\ln n$, with m' being the distance of the new gap to the closest edge of the Pareto front. To do so, we first perform centered crossover on $[a..b]$ and then create a segment of size at least ℓ from the element created by crossover by covering an interval of size ℓ.

Let $[a'..c], [d..b']$ be ℓ-gaps after performing the two steps, with $[c+1\,..\,d-1]$ a segment. Note that if more than two gaps are created by other operations, none of them can be larger than these two gaps. Of those two gaps, regard $[a'..c]$. Let its distance to the closest edge of the Pareto front be $m' = \frac{a'+c}{2}$.

We have $a' \geq a$. By definition of centered crossover, there is an $s \in [-\frac{1}{4}, \frac{1}{4}]$ such that $c = m + sw$. Then, $m' = m + \frac{sw}{2} \geq m - \frac{w}{8}$. With $w \leq (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n$, we get

$$m' \geq m - \frac{\left(\frac{3}{4}\right)^{\frac{i}{2}}\sqrt{m}\ln n}{8} \geq m - \frac{\sqrt{m}\ln n}{8}.$$

With $\ell \in \Omega(n^\delta) \cap o(n)$ for $\delta \in (0,1)$ and $m > \ell$, we get $\ln n < \sqrt{m}$ for sufficiently large n, so we have $m' \geq m - \frac{m}{8} = \frac{7}{8}m$.

Let the size of $[a'..c]$ be $w' = \frac{w}{2} + sw = (s + \frac{1}{2})w$.

With $w \leq (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n$, we get

$$w' \leq \left(s + \frac{1}{2}\right)\left(\frac{3}{4}\right)^{\frac{i}{2}}\sqrt{m}\ln n.$$

With $s \in [-\frac{1}{4}, \frac{1}{4}]$, this is at most

$$\frac{3}{4}\left(\frac{3}{4}\right)^{\frac{i}{2}}\sqrt{m}\ln n = \sqrt{\frac{3}{4}}\left(\frac{3}{4}\right)^{\frac{i}{2}}\sqrt{\frac{3}{4}m}\ln n = \left(\frac{3}{4}\right)^{\frac{i+1}{2}}\sqrt{\frac{3}{4}m}\ln n \leq \left(\frac{3}{4}\right)^{\frac{i+1}{2}}\sqrt{\frac{7}{8}m}\ln n$$
$$\leq \left(\frac{3}{4}\right)^{\frac{i+1}{2}}\sqrt{m'}\ln n.$$

We have therefore shown that $w' \leq (\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m'}\ln n$ and thereby that centered crossover and then covering a segment from the individual decreases the size of all created gaps sufficiently.

By Lemma 18, using $q = (\frac{3}{4})^{\frac{i}{2}}$, performing centered crossover takes an expected $\frac{160}{p_c}\sqrt{3}\big(\frac{3}{4}\big)^{\frac{i}{2}} \cdot \frac{n^2}{\ell^2}\ln^2 n$ objective function evaluations with probability in $1 - o(n^{-3})$. Performed in parallel on up to n gaps, a union bound yields a probability of $1 - o(n^{-2})$ of finishing in that time. Creating segments then takes $4en\ell(1-p_c)^{-1}$ with probability in $1 - o(n^{-3})$, using Lemma 5. Therefore, the total number of objective function evaluations is at most $\frac{160}{p_c}\sqrt{3}\big(\frac{3}{4}\big)^{\frac{i}{2}} \cdot \frac{n^2}{\ell^2}\ln^2 n + 4en\ell(1-p_c)^{-1}$ with probability in $1 - o(n^{-2})$. □

Applying Lemma 19 at most $O(\log n)$ times concludes the analysis of this phase.

Lemma 8 (Second Phase)**.** *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let P be a population such that $[0..\ell]$ and $[n-\ell\,..\,n]$ are covered intervals and that for all ℓ–gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, the gap has size at most $\max\{4\ell, \sqrt{m}\ln n\}$. Then there exists a constant $\alpha > 0$ such that it takes at most $\alpha(\frac{n^2}{\ell^2}\log^2 n + n\ell\log n)$ iterations/function evaluations with probability $1 - o(n^{-1})$ until all ℓ–gaps have size at most 4ℓ.*

Proof. By Lemma 19, we know that it takes at most $\frac{160}{p_c}\sqrt{3}(\frac{3}{4})^{\frac{i}{2}}\frac{n^2}{\ell^2}\ln^2 n + 4en\ell(1-p_c)^{-1}$ objective function evaluations with probability in $1 - o(n^{-2})$ to reduce the size of all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ from at most $\max\{4\ell, (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n\}$ to at most $\max\{4\ell, (\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m}\ln n\}$. This gives us a runtime on the order of $(\frac{3}{4})^{\frac{i}{2}}\frac{n^2}{\ell^2}\log^2 n + n\ell$.

Repeated application, starting with a maximum gap size of $(\frac{3}{4})^{\frac{0}{2}}\sqrt{m}\ln n$ for m the distance of a given gap's midpoint the the edge of the Pareto front, gives us a maximum ℓ-gap size of $(\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n$ after i iterations. With this, we must perform both steps at most $\gamma\ln n$ times for a constant $\gamma > 0$ until the largest ℓ-gap has size less than 4ℓ. Summed up over $\gamma\ln n$ iterations, we arrive at a runtime on the order of

$$\sum_{i=1}^{\gamma\log n}\left(\left(\frac{3}{4}\right)^{\frac{i}{2}}\frac{n^2}{\ell^2}\log^2 n + n\ell\right) = \sum_{i=1}^{\gamma\log n}\left(\frac{3}{4}\right)^{\frac{i}{2}}\frac{n^2}{\ell^2}\log^2 n + \sum_{i=1}^{\gamma\log n} n\ell$$
$$= \frac{n^2}{\ell^2}\log^2 n \sum_{i=1}^{\gamma\log n}\left(\frac{3}{4}\right)^{\frac{i}{2}} + \gamma n\ell\log n.$$

Using the geometric series, we know that

$$\sum_{i=1}^{\gamma\ln n}\left(\frac{3}{4}\right)^{\frac{i}{2}} \leq \frac{1}{1-\big(\frac{3}{4}\big)^{\frac{1}{2}}},$$

which is constant. Therefore, the total runtime is on the order of $\frac{n^2}{\ell^2}\log^2 n + n\ell\log n$. The probability for this is at least $1 - o(n^{-1})$, given by a union bound over less than n applications of Lemma 19. □

A.5 Material for Section 5

For our proof, we first bound the number of iterations it takes to cover the intervals using mutation, which can happen simultaneously for multiple intervals at once. This proof is largely analogous to the one given in Lemma 5 for the (G)SEMO.

Lemma 20. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let a p_c be the constant crossover rate. Let $x_1, \ldots, x_k$ be distinct elements of a population P and $J_1, \ldots, J_k$ intervals, where, for $i \in [k]$, $J_i = [x_i\,..\,x_i + \ell]$ if $|x_i|_1 \leq \frac{n}{2}$ and $J_i = [x_i - \ell\,..\,x_i]$ if $|x_i|_1 > \frac{n}{2}$. Then the number of iterations it takes to cover all intervals $J_1, \ldots, J_k$ is at most $4e\ell((1-p_c)(1-e^{-1}))^{-1}$ with probability $1 - o(n^{-2})$.*

Proof. Fix an interval J_i and let x be an individual with $|x|_1 \in J_i$. By symmetry, we may assume that $|x|_1 \leq \frac{n}{2}$. Then $|x|_1 \geq \frac{n}{2} - \ell$. For sufficiently large n, $\ell < \frac{n}{6}$. It follows that $|x|_0 \geq \frac{n}{3}$, so the probability of flipping such a bit is at least $\frac{1}{3}$, conditioned on choosing mutation and then flipping exactly one bit, which happens with probability at least $\frac{1}{e}(1-p_c)$. Selecting a specific individual during offspring generation at least once has probability $1-(1-N^{-1})^N \geq 1-e^{-1}$.

Pessimistically assuming that we cover the interval sequentially, the iterations needed to cover it is stochastically dominated by a sum of ℓ geometric random variables with success probabilities $\frac{1}{3e}(1-p_c)(1-e^{-1})$.

Denote these variables by $G_1, \dots, G_\ell$ and their sum by $X_i \coloneqq \sum_{j=1}^{\ell} G_j$. Then we have $\mathrm{E}[X_i] = 3e\ell((1-p_c)(1-e^{-1}))^{-1}$. We may consider the intervals independently, as any influence these processes have on each other will not worsen the runtime.

By Theorem 12 we have

$$\begin{aligned}\Pr\left[X_i \geq \left(1+\frac{1}{3}\right)\cdot 3e\ell((1-p_c)(1-e^{-1}))^{-1}\right] &\leq \exp\left(-\frac{1}{36}(\ell-1)\right)\\ &\leq e\exp\left(-\frac{1}{36}\ell\right)\\ &\in o(n^{-3}).\end{aligned}$$

The event that any of these intervals are not covered within $4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations is the union of any single one not being covered in that time. Therefore, by the union bound, it has probability at most $k \cdot o(n^{-3}) \subseteq n \cdot o(n^{-3}) = o(n^{-2})$. □

A.6 First Phase

Recall the definition of centered crossover (Definition 6) and the probability of a single crossover operation being centered (Lemma 15), both of which hold for the NSGA-II as well. We now prove a bound on the number of iterations until the maximum gap size is decreased by a constant factor, similar to the one given in Lemma 16.

Lemma 21. *Let $w \in \mathbb{N}$ and $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$. Let P be a population such that all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, whose size is at least $\max\{4\ell, \sqrt{m}\ln n\}$, have size at most w. Let T be the number of iterations of the NSGA-II with uniform crossover and constant crossover rate p_c until all these gaps have size at most $\frac{3}{4}w$.*

Then, $\Pr[T \leq 4p_c^{-1}\frac{N}{\ell^2}\ln n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}] \in 1-o(n^{-2})$.

Proof. Let $[a..b]$ be an ℓ-gap of size at most w and $m = \min\{\frac{a+b}{2}, n-\frac{a+b}{2}\}$. Due to symmetry, we may assume that $m = \frac{a+b}{2}$. Since we are not interested in gaps of size less than $\max\{4\ell, \sqrt{m}\ln n\}$, assume that $w > \max\{4\ell, \sqrt{m}\ln n\}$.

We are now interested in the time it takes to split $[a..b]$ into at least two gaps, each of size at most $\frac{3}{4}w$.

First, note that performing centered crossover on a gap $[a..b]$, yielding a new individual x with $c \coloneqq |x|_1 \in [m-\frac{w}{4}, m+\frac{w}{4}]$, we obtain two intervals $[a\,..\,c-1]$ and $[c+1\,..\,b]$, neither of which containing a segment of size ℓ and each of size at most $\frac{3}{4}w$. If this centered crossover is followed by either covering the interval $[c\,..\,c+\ell]$ or the interval $[c-\ell\,..\,c]$, we obtain at least two gaps, each of size at most $\frac{3}{4}w$.

Regarding the time for centered crossover, let $[a_1..b_1], \dots, [a_k..b_k]$ be the gaps at a given time. We know each gap to be bordered by segments of size at least ℓ by definition. Therefore, for each gap, the probability of choosing one parent each from the ℓ adjacent elements to either side of the gap is at least $p_c\frac{\ell^2}{N^2}$ per offspring generation. The probability of selecting such a pair of individuals during the N offspring generations in one iteration at least once is $1-(1-p_c\frac{\ell^2}{N^2})^N \geq 1-\exp(-p_c\frac{\ell^2}{N})$. Thus, the probability that we need to wait more than $4p_c^{-1}\frac{N}{\ell^2}\ln n$ iterations until a given gap has produced

offspring by crossover is

$$\left(1-\left(1-\exp\left(-p_c\frac{\ell^2}{N}\right)\right)\right)^{4p_c^{-1}\frac{N}{\ell^2}\ln n} \leq \exp(-4\ln n) \in o(n^{-3})$$

According to Lemma 15, the probability that that crossover is not centered is at most $2 \cdot n^{-\frac{1}{128}\ln n}$. Therefore, the probability, that no centered crossover has been performed on the gap in the given time is in $o(n^{-3})$. A union bound over all $k \leq n$ gaps yields a probability in $o(n^{-2})$ of this step taking more than $4p_c^{-1}\frac{N}{\ell^2}\ln n$ iterations.

Lemma 20 bounds the number of iterations needed for covering intervals of size ℓ to be on the order of $4e\ell((1-p_c)(1-e^{-1})^{-1}$ with probability $1-o(n^{-2})$. Taken together, reducing the size of all gaps to at most $\frac{3}{4}w$ takes at most $4p_c^{-1}\frac{N}{\ell^2}\ln n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations with probability $1-o(n^{-2})$. □

With this, we can now bound the runtime of the NSGA-II until the maximum gap size is too small for the previous crossover strategy to succeed with sufficiently high probability.

Lemma 22 (First Phase of the NSGA-II)**.** *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let P be a population of the NSGA-II with uniform crossover and constant crossover probability p_c such that the intervals $[0..\ell]$ and $[n-\ell..n]$ are covered. Then there exists a constant $\alpha > 0$ such that it takes at most $\alpha(\frac{N}{\ell^2}\log^2 n + \ell\log n)$ iterations ($\alpha(\frac{N^2}{\ell^2}\log^2 n + N\ell\log n)$ objective function evaluations) with probability $1-o(n^{-1})$ until all ℓ–gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n-\frac{a+b}{2}\}$ are of size at most $\max\{4\ell, \sqrt{m}\ln n\}$.*

Proof. We can apply Lemma 21 to reduce the maximum ℓ–gap size by a factor of $\frac{3}{4}$. Each of those reductions takes at most $4p_c^{-1}\frac{N}{\ell^2}\ln n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations with probability $1-o(n^{-2})$.

After applying Lemma 21 on the order of $\ln n$ times, all gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n-\frac{a+b}{2}\}$ have size at most $\max\{4\ell, \sqrt{m}\ln n\}$. A union bound over the failure probabilities gives us a probability of $1-o(n^{-1})$ of finishing within $O(\frac{N}{\ell^2}\log^2 n + \ell\log n)$ iterations. □

A.7 Second Phase

From there we have to switch the strategy for crossover so that we may still decrease the maximum gap size by a constant factor, using the same argument as in the second phase for the (G)SEMO. For this we again aim to perform crossover with a larger number of individuals, which need not be part of an adjacent segment of a gap. The probability given in Lemma 17 for the (G)SEMO likewise holds for the NSGA-II, so we can proceed directly to proving a bound on the expected number of iterations until a single crossover attempt for appropriately chosen individuals succeeds. This proof is similar to the one in Lemma 18.

Lemma 23. *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$ and let $[a..b]$ be an ℓ-gap. Further, let $q \leq 1$ and, for all ℓ-gaps $[e..f]$ with size w_{ef} and $m_{ef} = \min\{\frac{e+f}{2}, n-\frac{e+f}{2}\}$, assume $4\ell \leq w_{ef} \leq q\sqrt{m_{ef}}\ln n$. Then, with a constant crossover rate p_c, it takes at most $160\sqrt{3}\cdot q\cdot\frac{N}{p_c\ell^2}\ln^2 n$ iterations ($160\sqrt{3}\cdot q\cdot\frac{N^2}{p_c\ell^2}\ln^2 n$ objective function evaluations) with probability $1-o(n^{-3})$ until centered crossover is performed on $[a..b]$.*

Proof. Let $m = \min\{\frac{a-b}{2}, n-\frac{a-b}{2}\}$. By symmetry, we may assume that $m = \frac{a+b}{2}$, meaning the distance to the closest edge of the Pareto front corresponds to the number of 1-bits of the midpoint.

We first want to choose individuals x and y to perform crossover on. We start by regarding the number of choices we have for each.

First, we want to choose an individual y from the interval $[b+\frac{15}{16}m\,..\,b+m]$. We distinguish two cases. For the first one, assume an ℓ-gap with its midpoint in $[b+\frac{15m}{16}..b+m]$ exists. Take any such

ℓ-gap $[e..f]$. Then, the size of the gap $[e..f]$ is at most

$$q\sqrt{\frac{e+f}{2}}\ln n \leq q\sqrt{\frac{2(b+m)}{2}}\ln n = q\sqrt{b+m}\ln n \leq q\sqrt{3m}\ln n,$$

since $b \leq 2m$. We know that segments $[e' \,..\, e-1], [f+1 \,..\, f']$ of size at least ℓ must exist. Therefore, the existence of any ℓ-gap implies the existence of one segment of size at least ℓ. For a lower bound on the number of covered elements, assume that each of those segments has size exactly ℓ. Regarded over the $\frac{m}{16}$ elements in $[b+\frac{15m}{16} \,..\, b+m]$, this gives us at least $\ell \cdot \frac{\frac{m}{16}}{q\sqrt{3m}\ln n+\ell}$ individuals. With $\sqrt{3}\cdot q\ln n > 4\ell$, this is greater than $\ell \cdot \frac{\frac{m}{16}}{2q\sqrt{3m}\ln n} = \ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q\ln n}$. With this, we get at least $\ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q\ln n}$ covered elements in $[b+\frac{15m}{16} \,..\, b+m]$ to choose y from.

For the second case, assume no ℓ-gap $[e..f]$ with $\frac{e+f}{2} \in [b+\frac{15m}{16} \,..\, b+m]$ exists. Let $[g..h]$ be any gap such that $[b+\frac{15m}{16} \,..\, b+m] \cap [g..h] \neq \emptyset$ and let $m_{gh} = \min(\frac{g+h}{2}, n-\frac{g+h}{2}) \leq \frac{g+h}{2}$. Then, $g \leq b+\frac{15m}{16}$. Let w_{gh} be the size of $[g..h]$. Then, we know that $m_{gh} \leq \frac{g+h}{2} = \frac{2g+w_{gh}}{2} \leq b+\frac{15m}{16}+\frac{w_{gh}}{2}$. With $w_{gh} \leq q\sqrt{m_{gh}}\ln n$, we have $w_{gh} \leq m_{gh}$ for sufficiently large n and with $m_{gh} > \ell$. With this and $b \leq 2m$, we have $m_{gh} \leq b+\frac{15m}{16}+\frac{w_{gh}}{2} \leq 2m+\frac{15m}{16}+\frac{m_{gh}}{2}$, so $m_{gh} \leq 2(2m+\frac{15m}{16})$. Accordingly, we get $w_{gh} \leq q\sqrt{2(2m+\frac{15m}{16})}\ln n$. For sufficiently large n, this is less than $\frac{1}{4}$ of $[b+\frac{15m}{16} \,..\, b+m]$. Therefore, with at most two such gaps (one from either side of the interval), at most half of $[b+\frac{15m}{16} \,..\, b+m]$ is part of an ℓ-gap for sufficiently large n, giving us at least $\frac{1}{2}\frac{m}{16}$ elements to choose from for y. Combined with the other case and with $m > \ell$, we get at least $\ell \cdot \frac{\sqrt{m}}{32\sqrt{3}\cdot q\ln n}$ elements to choose from for y.

Let $y \in P$ with $|y|_1 \in [b+\frac{15m}{16} \,..\, b+m]$. Let $d \in \mathbb{N}$ be the distance of y to m, meaning $|y|_1 = m+d$. For choosing x, we want at least ℓ possible elements with the number of 1-bits as close to $m-d$ as possible. However, it is possible for $m-d$ to be part of a gap. Call that gap $[g..h] \subseteq [0 \,..\, a-\frac{15m}{16}]$. Its size is then at most $q\ln n\sqrt{\frac{g+h}{2}} \leq q\ln n\sqrt{\frac{2(a-\frac{15m}{16})}{2}} \leq q\ln n\sqrt{\frac{m}{16}} = \frac{1}{4}q\ln n\sqrt{m}$. Therefore, for an $r \leq \frac{1}{4}q\ln n\sqrt{m}$, there exists a segment $[z \,..\, m-d-r]$ of size at least ℓ. We select x from the elements in $[m-d-r-\ell \,..\, m-d-r]$ of this segment. With this, we have at least $\ell \cdot \frac{\sqrt{m}}{2\sqrt{3}\cdot q\ln n}$ choices for selecting x and at least ℓ choices for selecting y.

Selecting any such pair for crossover during one iteration during the generation of the offspring at least once has probability

$$1-\left(1-\frac{p_c\ell^2\sqrt{m}}{N^2\cdot 32\sqrt{3}\cdot q\ln n}\right)^N \geq 1-\exp\left(-\frac{p_c\ell^2\sqrt{m}}{N\cdot 32\sqrt{3}\cdot q\ln n}\right).$$

With this, the probability that any such pair has not been selected for crossover in the first $160\sqrt{3}\cdot q\cdot\frac{N}{p_c\ell^2\sqrt{m}}\ln^2 n$ steps is

$$\left(1-\left(1-\exp\left(-\frac{p_c\ell^2\sqrt{m}}{N\cdot 32\sqrt{3}\cdot q\ln n}\right)\right)\right)^{160\sqrt{3}\cdot q\cdot\frac{N}{p_c\ell^2\sqrt{m}}\ln^2 n} \leq \exp(-5\ln n) \in o\big(n^{-4}\big).$$

By Lemma 17, we have a probability of at least $\beta(2m)^{-\frac{1}{2}}\sqrt{\ell}$ of the performed crossover operation being centered on $[a..b]$. Performing $\sqrt{m}$ crossover operations, the probability of none of them being centered is

$$\left(1-\frac{\beta\sqrt{\ell}}{\sqrt{2m}}\right)^{\sqrt{m}} \leq \exp\left(-\frac{\beta}{\sqrt{2}}\sqrt{\ell}\right).$$

This means we have a probability of at most $\exp(-\beta(\frac{\ell}{2})^{\frac{1}{2}})$ of not performing centered crossover in $\sqrt{m}$ crossover operations, conditioned on having selected parents x and y as described for each crossover. A union bound over $\sqrt{m} \leq n$ crossover operations gives a probability of the parent

selection for all crossover operations combined taking longer than $160\sqrt{3} \cdot \sqrt{m} \cdot q \cdot \frac{N}{p_c \ell^2 \sqrt{m}} \ln^2 n = 160\sqrt{3} \cdot q \cdot \frac{N}{p_c \ell^2} \ln^2 n$ iterations in $o(\sqrt{m} n^{-4}) \subseteq o(n^{-3})$. With $\exp(-\beta(\frac{\ell}{2})^{\frac{1}{2}}) \in o(n^{-3})$, the probability of not performing centered crossover on $[a..b]$ in at most $160\sqrt{3} \cdot q \cdot \frac{N}{p_c \ell^2} \ln^2 n$ iterations is in $o(n^{-3})$. □

We now combine the results from Lemma 23 and Lemma 20, similar again to Lemma 19 for the (G)SEMO.

Lemma 24. *Let $i \in \mathbb{N}$ and let $\ell \in \Omega(n^\delta) \cap o(n)$ for a constant $\delta \in (0,1)$. Let P be a population such that for all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, the gap has size at most $\max\{4\ell, (\frac{3}{4})^{\frac{i}{2}} \sqrt{m} \ln n\}$. Then, the NSGA-II with uniform crossover and constant crossover rate p_c takes at most $\frac{160}{p_c}\sqrt{3}(\frac{3}{4})^{\frac{i}{2}} \cdot \frac{N}{\ell^2} \ln^2 n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations ($\frac{160}{p_c}\sqrt{3}(\frac{3}{4})^{\frac{i}{2}} \cdot \frac{N^2}{\ell^2} \ln^2 n + 4e\ell N((1-p_c)(1-e^{-1}))^{-1}$ objective function evaluations) with probability $1 - o(n^{-1})$ until all ℓ-gaps have size at most $\max\{4\ell, (\frac{3}{4})^{\frac{i+1}{2}} \sqrt{m} \ln n\}$.*

Proof. Let $[a..b]$ be an ℓ-gap with size w and $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Due to symmetry, we may assume that $m = \frac{a+b}{2}$. Since we are only interested in ℓ-gaps of size greater than 4ℓ, assume that $4\ell < w \leq (\frac{3}{4})^{\frac{i}{2}} \sqrt{m} \ln n$. Likewise, assume that $w \geq (\frac{3}{4})^{\frac{i+1}{2}} \sqrt{m} \ln n$, as otherwise, $[a..b]$ already fulfills the condition.

We are now interested in the time it takes to split $[a..b]$ into at least two gaps, each of size at most $(\frac{3}{4})^{\frac{i+1}{2}} \sqrt{m'} \ln n$, with m' being the distance of the new gap to the closest edge of the Pareto front. To do so, we first perform centered crossover on $[a..b]$ and then create a segment of size at least ℓ from the element created by crossover by covering an interval of size ℓ.

Let $[a'..c], [d..b']$ be ℓ-gaps after performing the two steps, with $[c+1 \,..\, d-1]$ a segment. Note that if more than two gaps are created by other operations, none of them can be larger than these two gaps. Of those two gaps, regard $[a'..c]$. Let its distance to the closest edge of the Pareto front be $m' = \frac{a'+c}{2}$.

We have $a' \geq a$. By definition of centered crossover, there is an $s \in [-\frac{1}{4}, \frac{1}{4}]$ such that $c = m + sw$. Then, $m' = m + \frac{sw}{2} \geq m - \frac{w}{8}$. With $w \leq (\frac{3}{4})^{\frac{i}{2}} \sqrt{m} \ln n$, we get

$$m' \geq m - \frac{\left(\frac{3}{4}\right)^{\frac{i}{2}} \sqrt{m} \ln n}{8} \geq m - \frac{\sqrt{m} \ln n}{8}.$$

With $\ell \in \Omega(n^\delta) \cap o(n)$ for $\delta \in (0,1)$ and $m > \ell$, we get $\ln n < \sqrt{m}$ for sufficiently large n, so we have $m' \geq m - \frac{m}{8} = \frac{7}{8}m$.

Let the size of $[a'..c]$ be $w' = \frac{w}{2} + sw = (s + \frac{1}{2})w$.

With $w \leq (\frac{3}{4})^{\frac{i}{2}} \sqrt{m} \ln n$, we get

$$w' \leq \left(s + \frac{1}{2}\right)\left(\frac{3}{4}\right)^{\frac{i}{2}} \sqrt{m} \ln n.$$

With $s \in [-\frac{1}{4}, \frac{1}{4}]$, this is at most

$$\begin{aligned} \frac{3}{4}\left(\frac{3}{4}\right)^{\frac{i}{2}} \sqrt{m} \ln n &= \sqrt{\frac{3}{4}}\left(\frac{3}{4}\right)^{\frac{i}{2}} \sqrt{\frac{3}{4} m} \ln n = \left(\frac{3}{4}\right)^{\frac{i+1}{2}} \sqrt{\frac{3}{4} m} \ln n \leq \left(\frac{3}{4}\right)^{\frac{i+1}{2}} \sqrt{\frac{7}{8} m} \ln n \\ &\leq \left(\frac{3}{4}\right)^{\frac{i+1}{2}} \sqrt{m'} \ln n. \end{aligned}$$

We have therefore shown that $w' \leq (\frac{3}{4})^{\frac{i+1}{2}} \sqrt{m'} \ln n$ and thereby that centered crossover and then covering a segment from the individual decreases the size of all created gaps sufficiently.

By Lemma 23, using $q = (\frac{3}{4})^{\frac{i}{2}}$, performing centered crossover takes an expected $\frac{160}{p_c}\sqrt{3}\big(\frac{3}{4}\big)^{\frac{i}{2}} \cdot \frac{N}{\ell^2}\ln^2 n$ iterations with probability in $1 - o(n^{-3})$. Performed in parallel on up to n gaps, a union bound yields a probability of $1 - o(n^{-2})$ of finishing in that time. Creating segments then takes $4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations with probability in $1 - o(n^{-2})$, using Lemma 20. Therefore, the total number of iterations is at most $\frac{160}{p_c}\sqrt{3}\big(\frac{3}{4}\big)^{\frac{i}{2}} \cdot \frac{N}{\ell^2}\ln^2 n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}$ with probability in $1 - o(n^{-2})$. □

With this, we can complete the proof of the second phase, that is, we prove a bound on the number of iterations until the maximum gap size is at most 4ℓ with a proof similar to the one in Lemma 8 for the (G)SEMO.

Lemma 25 (Second Phase of the NSGA-II)**.** *Let $\ell \in \Omega(n^\delta) \cap o(n)$ for some $\delta \in (0,1)$ and let P be a population such that $[0..\ell]$ and $[n-\ell\,..\,n]$ are covered intervals and that for all ℓ–gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$, the gap has size at most $\max\{4\ell, \sqrt{m}\ln n\}$. Consider the NSGA-II with uniform crossover and constant crossover rate p_c. Then there exists a constant $\alpha > 0$ such that it takes at most $\alpha(\frac{N}{\ell^2}\log^2 n + \ell\log n)$ iterations ($\alpha(\frac{N^2}{\ell^2}\log^2 n + N\ell\log n)$ objective function evaluations) with probability $1 - o(n^{-1})$ until all ℓ–gaps have size at most 4ℓ.*

Proof. By Lemma 24, we know that it takes at most $\frac{160}{p_c}\sqrt{3}\big(\frac{3}{4}\big)^{\frac{i}{2}} \cdot \frac{N}{\ell^2}\ln^2 n + 4e\ell((1-p_c)(1-e^{-1}))^{-1}$ iterations with probability in $1-o(n^{-2})$ to reduce the size of all ℓ-gaps $[a..b]$ with $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$ from at most $\max\{4\ell, (\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n\}$ to at most $\max\{4\ell, (\frac{3}{4})^{\frac{i+1}{2}}\sqrt{m}\ln n\}$.

Repeated application, starting with a maximum gap size of $(\frac{3}{4})^{\frac{0}{2}}\sqrt{m}\ln n$ for m the distance of a given gap's midpoint the the edge of the Pareto front, gives us a maximum ℓ-gap size of $(\frac{3}{4})^{\frac{i}{2}}\sqrt{m}\ln n$ after i iterations. With this, we must perform both steps at most $\gamma\ln n$ times for a constant $\gamma > 0$ until the largest ℓ-gap has size less than 4ℓ. Summed up over $\gamma\ln n$ iterations, we arrive at a runtime on the order of

$$\sum_{i=1}^{\gamma\log n}\left(\left(\frac{3}{4}\right)^{\frac{i}{2}}\frac{N}{\ell^2}\log^2 n + \ell\right) = \sum_{i=1}^{\gamma\log n}\left(\frac{3}{4}\right)^{\frac{i}{2}}\frac{N}{\ell^2}\log^2 n + \sum_{i=1}^{\gamma\log n}\ell$$
$$= \frac{N}{\ell^2}\log^2 n\sum_{i=1}^{\gamma\log n}\left(\frac{3}{4}\right)^{\frac{i}{2}} + \gamma\ell\log n.$$

Using the geometric series, we know that

$$\sum_{i=1}^{\gamma\ln n}\left(\frac{3}{4}\right)^{\frac{i}{2}} \le \frac{1}{1-\big(\frac{3}{4}\big)^{\frac{1}{2}}},$$

which is constant. Therefore, the total number of iterations is on the order of $\frac{N}{\ell^2}\log^2 n$. The probability for this is at least $1 - o(n^{-1})$, given by a union bound over less than n applications of Lemma 24. □

A.8 Final Proof

Combining the bounds for the first and the second phase, and again using Lemma 20, we now prove the main theorem of this section.

Proof of Theorem 9. Define $\ell := \lceil\sqrt[3]{N\ln n}\rceil$. With Lemma 20, we can cover the intervals $[0..\ell]$ and $[n-\ell\,..\,n]$ in $O(\ell) = O(N^{1/3}\log^{1/3} n)$ iterations with probability $1 - o(n^{-2})$. With this, we have at least one ℓ-gap. Lemma 22 then yields that in time $O(NN^{-2/3}\log^2 n\log^{-2/3} n + N^{1/3}\log n\log^{1/3} n) = O(N^{1/3}\log^{4/3} n)$ with probability $1 - o(n^{-1})$ all ℓ-gaps $[a..b]$ have size at most $\max\{4\ell, \sqrt{m}\log n\}$, where $m = \min\{\frac{a+b}{2}, n - \frac{a+b}{2}\}$. Lemma 25 continues from there and gives that in time $O(N^{1/3}\log^{4/3} n)$ with probability $1 - o(n^{-1})$ all ℓ-gaps have size at most 4ℓ. To

cover those ℓ-gaps, we can apply Lemma 20 again, leading to an additional time of $O(N^{1/3}\log^{1/3} n)$ with probability $1-o(n^{-2})$. This covers the Pareto front. In total, there exists a constant $\alpha > 0$ such that $T < \alpha N^{1/3}\log^{4/3} n$ with probability $1-o(n^{-1})$, proving the high-probability statement.

For the bound on the expectation, we can repeat the above argument, exploiting that our analysis applies to any initial state of the algorithm (that has the single-objective optima in the population). As a shorthand, define $t \coloneqq \alpha N^{1/3}\log^{4/3} n$. If we are not finished in the first t iterations, we can simply start the analysis anew with none of the individuals found previously worsening the runtime. Therefore, for all $i \geq 1$, we have $\Pr[T \geq it] \leq n^{-i}$. Now define $X \coloneqq \lceil \frac{T}{t} \rceil$. Clearly, $T \leq tX$. Since X only takes on positive integers, we can apply Lemma 14, yielding

$$\mathrm{E}[X] = \sum_{i=1}^{\infty} \Pr[X \geq i] \leq \sum_{i=1}^{\infty} \Pr[T \geq it] \leq \sum_{i=1}^{\infty} n^{-i} \leq \frac{1}{1-\frac{1}{n}} \leq 2.$$

Using $\mathrm{E}[T] \leq t\mathrm{E}[X]$, we get $\mathrm{E}[T] \leq 2t$, proving the proposition. □

A.9 Large Population Size

Lemma 10. *Consider the NSGA-II with population size $N \geq 4(n+1)$ with uniform crossover, and crossover rate p_c such that p_c and $1-p_c$ are $\Omega(1)$. Assume that the initial population is arbitrary, but contains at least one optimum of each objective. Assume we optimize the* OneMinMax *or* COCZ *benchmark defined on individuals of length n. Then, it takes $O(n)$ iterations ($O(Nn)$ function evaluations) both in expectation and with probability at least $1-o(n^{-1})$ until the population of the algorithm witnesses the Pareto front.*

Proof. By symmetry, consider the interval $L \coloneqq [0..\frac{n}{2}]$. and let x be an individual with $|x|_1 \in L$. Then $|x|_0 \geq \frac{n}{2}$, so the probability of flipping such a bit is at least $\frac{1}{2}$ conditioned on choosing mutation and then flipping exactly one bit. For the NSGA-II, this happens with probability at least $\frac{1}{e}(1-p_c)$. Selecting a specific individual during offspring generation at least once has probability $1-(1-N^{-1})^N \geq 1-e^{-1}$.

Pessimistically assuming that we cover the interval sequentially, starting by choosing the individual with $|x|_1 = 0$, the time to cover it is stochastically dominated by a sum of $\frac{n}{2}+1$ geometric random variables with success probabilities $\frac{1}{2e}(1-p_c)(1-e^{-1})$.

Denote these variables by $G_1, \dots, G_{\frac{n}{2}+1}$ and their sum by $X_L \coloneqq \sum_{j=1}^{n/2+1}$. Then we have $\mathrm{E}[X_L] = en((1-p_c)(1-e^{-1})^{-1}$. We may consider this interval and $[\frac{n}{2}..n]$ independently, as any influence these processes have on each other will not worsen the runtime.

By Theorem 12 we have

$$\Pr\left[X_L \geq 2en((1-p_c)(1-e^{-1}))^{-1}\right] \leq \exp\left(-\frac{1}{4}\cdot\frac{n}{2}\right) \in o(n^{-2}).$$

With a union bound over the two intervals, the probability that one of the intervals is not covered after $2en((1-p_c)(1-e^{-1}))^{-1}$ is at least $1-o(n^{-1})$. □