Towards a Rigorous Understanding of the Population Dynamics of the NSGA-III: Tight Runtime Bounds

Andre Opris

Chair of Algorithms for Intelligent Systems, University of Passau, Passau, Germany
andre.opris@uni-passau.de

Abstract

Evolutionary algorithms are widely used for multi-objective optimization, with NSGA-III being particularly effective for problems with more than three objectives, unlike NSGA-II. Despite its empirical success, its theoretical understanding remains limited, especially regarding runtime analysis. A central open problem concerns its population dynamics, which involve controlling the maximum number of individuals sharing the same fitness value during the exploration process. In this paper, we make a significant step towards such an understanding by proving tight runtime bounds for NSGA-III on the bi-objective OneMinMax (2-OMM) problem. We show that, for population sizes $n + 1 \leq \mu = O(\log(n)^c(n + 1))$ where $c < 1$ is a constant, NSGA-III requires $\Omega(n^2 \log n/\mu)$ generations in expectation for covering the Pareto front, providing one of the first lower bounds for NSGA-III on a classical benchmark. Complementing this, we also improve the best known upper bound for NSGA-III on the m-objective OneMinMax problem (m-OMM) of $O(n \log(n))$ generations by a factor of $\mu/(2n/m + 1)^{m/2}$ for constant m and $(2n/m + 1)^{m/2} \leq \mu \in O(\sqrt{\log n}(2n/m + 1)^{m/2})$. This yields tight runtime bounds for $m = 2$, and the surprising result that NSGA-III outperforms NSGA-II by a factor of μ/n in the expected runtime.

Introduction

Decision making is a fundamental aspect of many areas in artificial intelligence, where it is often important to explore trade-offs and compromises between different options before reaching a conclusion (Luukkonen et al. 2023). Such situations are often formulated as multi-objective optimization problems, which are typically tackled using evolutionary multi-objective algorithms (Stewart, Palmer, and DuPont 2021). They apply principles of nature to optimize functions with conflicting objectives, aiming to find a diverse Pareto-optimal set of solutions. This offers decision makers a range of trade-off solutions, enabling them to select the one that best aligns with their preferences (Tamssaouet et al. 2022). It is therefore not surprising that such algorithms have become essential tools, widely applied across various practical domains. These include artificial intelligence (Koziel, Pietrenko-Dabrowska, and Szczepanski 2025), often in combination with bioinformatics (Handl, Kell, and Knowles 2007; Mukhopadhyay et al. 2024), as well as constraint optimization (García et al. 2021), machine learning (Qu et al. 2021; Schneider, Bischl, and Thomas 2023), and engineering (Andersson 2003; Amamou and Jebari 2023). In particular, many of such real-world applications involve optimization problems with many objectives. However, a huge challenge is that, as the number of objectives increases, the Pareto front expands exponentially, making the problems increasingly complex. Additionally, identifying dependencies between individual objectives becomes more difficult. There are already differences between two and more objectives. In the case of two objectives, sorting non-dominated individuals according to the first objective naturally leads to a reverse sorting with respect to the second, making the *crowding distance*, which measures the proximity of search points based on their sorting across objectives, a reliable indicator of their relative closeness. However, this relationship breaks down for problems with three or more objectives, as a solution can have a crowding distance of zero even when it is not close to other solutions (see, for example (Zheng and Doerr 2024a)). As a result, NSGA-II (Deb et al. 2002), the most cited EMOA ($\sim$ 56000 citations) which uses the crowding distance as a tie-breaker, succeeds in solving bi-objective problems (see for example (Zheng, Liu, and Doerr 2022) for a rigorous analysis or (Deb et al. 2002) for empirical results), but fails in optimizing many problems where the number of objectives is large (compare with (Zheng and Doerr 2024a) for large differences already between two and three objectives or for empirical studies (Khare, Yao, and Deb 2003)). To overcome this problem, Deb and Jain (2014) designed the NSGA-III algorithm. It uses a set of predefined reference points instead of the crowding distance. A major advantage is that these reference points can be predefined by users based on their specific needs. Hence, this algorithm has a huge practical impact ($\sim$6000 citations) and it is empirically shown that it can efficiently solve problems with at least four objectives (Deb and Jain 2014; Mkaouer et al. 2015; Yadav, Ramu, and Deb 2023). However, theoretical understanding of its success lags far behind its practical impact and the first papers addressing rigorous runtime analyses of this algorithm appeared only recently (see for example (Wietheger and Doerr 2023; Opris et al. 2024) for breakthroughs). Surprisingly, even in simple settings, Opris (2025a) showed that NSGA-III exhibits population dynamics that differ significantly from those of NSGA-II: NSGA-III successively iterates through all reference points, always choosing a point associated to a reference point with the fewest chosen individuals so far for the next generation, while NSGA-II treats all points with zero crowding distance equally. Hence, NSGA-III tends to spread solutions very evenly across the Pareto front (compare also with (Chaudhari et al. 2022) for empirical results). Indeed, for appropriate population sizes, it was shown in (Opris 2025a) that NSGA-III outperforms NSGA-II on the pseudo-Boolean bi-objective multimodal function ONEJUMPZEROJUMP (OJZJ for short) for appropriate population sizes. However, they showed how NSGA-III spreads solutions evenly across the Pareto front *after* converging to local optima. How this distribution evolves during exploration, particularly before reaching a local optimum or when no local optima exist at all, as in m-OMM, remains unclear. It is still unknown *when* and *why* NSGA-III performs well, or how quickly it spreads solutions across the Pareto front. As a first step toward understanding its limitations on many-objective problems, we focus on the bi-objective case, which already exhibits complex population dynamics.

Our contribution: We significantly increase the understanding of the population dynamics of the NSGA-III on the pseudo-Boolean 2-OMM by investigating the *maximum cover number* β, defined as the maximum number of individuals in the population sharing the same fitness vector, which is non-increasing as shown in (Opris 2025a). Our first two results are about the time to firstly cover a subset $\mathcal{A}$ of the Pareto front of a given cardinality α (Lemma 3), and then spread all solutions evenly on that set (Lemma 4). With high probability, this time is $O(\alpha)$. On the other hand, for a given maximum cover number β, we analyze the population's exploration towards the extreme point 1^n. Specifically, for two constants $0 < a < b \leq 3/4$, we provide a lower bound on the time required to reduce the minimum Hamming distance of an individual from the population to 1^n from at least n^b to at most n^a. With high probability, this time is $\Omega(n \ln n/\beta)$ (Lemmas 5 and 6). This bound increases asymptotically with $1/\beta$, which is unsurprising, since a smaller β reduces the probability of selecting individuals already close to 1^n, and further decreasing its distance to 1^n through mutation. Then, together with Lemmas 3 and 4, the cover number can be further reduced by first covering and then spreading the population over a set of larger cardinality $\Omega(n \ln n/\beta)$. This again results, by Lemmas 5 and 6, in a larger lower bound of $\Omega(n \ln n/\beta)$. By repeating this process a finite number of times, one can finally reduce β to a value of $O(\mu/n)$ before creating a search point x whose Hamming distance to 1^n is at most $n^{1/16}$. Notably, this is the smallest possible value for β up to a multiplicative constant, and leads to a lower bound of $\Omega(n^2 \ln(n)/\mu)$ expected generations for NSGA-III to optimize 2-OMM (Theorem 7). Finally, we improve the upper bound from (Opris et al. 2024) for any constant number m of objectives, where $(2n/m + 1)^{m/2} \leq \mu \in O(\sqrt{\ln(n)}(2n/m + 1)^{m/2})$, by a factor of $O((2n/m + 1)^{m/2}/\mu)$. In particular, we show that m-OMM can be optimized in $O\left(n \ln(n)(2n/m + 1)^{m/2}/\mu\right)$ generations in expectation (Theorem 8). This aligns with the earlier lower bound for $m = 2$ and shows the somewhat surprising result that NSGA-III outperforms NSGA-II, the state of the art algorithm for two objectives and widely used in practice (approximately 60,000 citations), by a factor of μ/n (see (Doerr and Qu 2023a) for a lower bound of $\Omega(n \ln n)$ generations for NSGA-II to optimize 2-OMM, assuming $4(n + 1) \leq \mu \leq o(n^\nu)(n + 1)$ for $\nu < 1$).

Related work: The mathematical runtime analysis of modern practical MOEAs has begun only recently. Zheng, Liu, and Doerr (2022) conducted the first runtime analysis of NSGA-II on classical benchmark functions, which inspired numerous follow-up studies on bi-objective optimization (Doerr and Qu 2022; Dang et al. 2023; Dang, Opris, and Sudholt 2025a, 2024a,b; Doerr and Qu 2023b; Zheng and Doerr 2022; Bian et al. 2023; Dang, Opris, and Sudholt 2025b), including extensions to combinatorial problems such as minimum spanning trees or subset selection (Cerf et al. 2023; Deng et al. 2024). Variants of NSGA-II have also been proposed to address its limitations on many-objective problems using tie-breaking rules (Doerr, Ivan, and Krejca 2025) or alternative crowding distance measures (Zheng, Gao, and Doerr 2024). Among the most prominent many-objective algorithms are SPEA2, SMS-EMOA, and NSGA-III, which have been successfully analyzed (Wietheger and Doerr 2023; Zheng and Doerr 2024b; Opris et al. 2024; Wietheger and Doerr 2024; Opris 2025c,a). Until (Opris 2025a; Doerr and Qu 2023a), there were no tight runtime bounds for NSGA-II or NSGA-III on classical benchmarks, despite extensive research on EMOA limitations and population dynamics. For instance, Opris (2025b) analyzed PAES-25 with one-bit mutation on the m-objective LEADINGONESTRAILINGZEROS problem, proving tight runtimes of $\Theta(n^3)$ for $m = 2$, $\Theta(n^3 \ln n)$ for $m = 4$, and $\Theta(n(2n/m)^{m/2} \ln(n/m))$ for $m > 4$. Additional tight bounds for GSEMO on several classical bi-objective benchmarks like OMM, OJZJ, and COUNTINGONESCOUNTINGZEROS are given in (Doerr, Krejca, and Opris 2025). To our knowledge, no results exist on the population dynamics of NSGA-III apart from investigations on OJZJ in (Opris 2025a).

Preliminaries

Given two random variables X and Y on $\mathbb{N}_0$, we say that Y stochastically dominates X if $\Pr(Y \leq c) \leq \Pr(X \leq c)$ for all $c \geq 0$. The number of ones in a bit string x is denoted by $|x|_1$ and the number of zeros by $|x|_0$, respectively. For any finite set A, we write $|A|$ to denote its cardinality. For $n \in \mathbb{N}$ let $[n] := \{1, \ldots, n\}$, denote by $\ln$ the natural logarithm (i.e. to base e) and let $\text{poly}(n)$ be a placeholder for some polynomial in n.

This paper is about many-objective optimization, specifically the maximization of a discrete m-objective function $f(x) := (f_1(x), \ldots, f_m(x))$ for $m \in \mathbb{N}$ where each $f_i : \{0,1\}^n \to \mathbb{N}_0$ for $i \in \{1, \ldots, m\}$. When $m = 2$, the function is also called *bi-objective*. For a bit string x let $x := (x^1, \ldots, x^{m/2})$ where all x^j are of equal length $2n/m$. For a subset $N \subseteq \{0,1\}^n$, we define $f(N) := \{f(x) \mid x \in N\}$. Given two search points $x, y \in \{0,1\}^n$, x *weakly dominates*

Algorithm 1: NSGA-III ((Deb and Jain 2014)) with population size μ on an m-objective function f

Initialize $P_0 \sim \text{Unif}((\{0,1\}^n)^\mu)$
for $t := 0$ *to* ∞ **do**
  Initialize $Q_t := \emptyset$
  for $i = 1$ *to* μ **do**
    Sample s from P_t uniformly at random
    Create r by standard bit mutation on s with mutation probability $1/n$
    Update $Q_t := Q_t \cup \{r\}$
  Set $R_t := P_t \cup Q_t$
  Partition R_t into layers $F_t^1, F_t^2, \dots, F_t^k$ of non-dominated solutions
  Find $i^* \geq 1$ such that $\sum_{i=1}^{i^*-1} |F_t^i| < \mu$ and $\sum_{i=1}^{i^*} |F_t^i| \geq \mu$
  Compute $Y_t = \bigcup_{i=1}^{i^*-1} F_t^i$
  Choose $\tilde{F}_t^{i^*} \subset F_t^{i^*}$ such that $|Y_t \cup \tilde{F}_t^{i^*}| = \mu$ with Algorithm 2
  Create the next population $P_{t+1} := Y_t \cup \tilde{F}_t^{i^*}$

Algorithm 2: Selection procedure utilizing a set $\mathcal{R}_p$ of reference points to maximize an m-objective function f

Compute the normalisation f^n of f
Associate each $x \in Y_t \cup F_t^{i^*}$ with its reference point $\text{rp}(x)$ such that the distance between $f^n(x)$ and the line through the origin and $\text{rp}(x)$ is minimized
For each $r \in \mathcal{R}_p$, set $\rho_r := |\{x \in Y_t \mid \text{rp}(x) = r\}|$
Initialize $\tilde{F}_t^{i^*} = \emptyset$ and $R' := \mathcal{R}_p$
while *true* **do**
  Determine $r_{\min} \in R'$ such that $\rho_{r_{\min}}$ is minimal (where ties are broken randomly)
  Determine $x_{r_{\min}} \in F_t^{i^*} \setminus \tilde{F}_t^{i^*}$ which is associated with $r_{\min}$ and minimizes the distance between the vectors $f^n(x_{r_{\min}})$ and $r_{\min}$ (where ties are broken randomly)
  if $x_{r_{\min}}$ *exists* **then**
    $\tilde{F}_t^{i^*} = \tilde{F}_t^{i^*} \cup \{x_{r_{\min}}\}$
    $\rho_{r_{\min}} = \rho_{r_{\min}} + 1$
    if $|Y_t| + |\tilde{F}_t^{i^*}| = \mu$ **then**
      return $\tilde{F}_t^{i^*}$
  else $R' = R' \setminus \{r_{\min}\}$;

y, denoted by $x \succeq y$, if $f_i(x) \geq f_i(y)$ for all $i \in [m]$ and x *(strictly) dominates* y, denoted by $x \succ y$, if one inequality is strict; if neither $x \succeq y$ nor $y \succeq x$ then x and y are *incomparable*. A set $S \subseteq \{0,1\}^n$ is a *set of mutually incomparable solutions* with respect to f if all search points in S are incomparable. Each solution not dominated by any other in $\{0,1\}^n$ is called *Pareto-optimal*. A mutually incomparable set of these solutions that covers all possible non-dominated fitness values is called a *Pareto(-optimal) set* of f. For a population $P_t \subset \{0,1\}^n$ and $v \in \mathbb{N}_0^m$ denote by $c_t(v) := |\{x \in P_t \mid f(x) = v\}|$ the *cover number* of v, i.e. the number of individuals from P_t with fitness vector v. We say that v is *covered* if $c_t(v) \geq 1$.

The NSGA-III algorithm, originated in (Deb and Jain 2014) is shown in Algorithm 1. Initially, a population of size μ is created by choosing μ individuals from $\{0,1\}^n$ uniformly at random. Then in each iteration t, a multiset Q_t of μ new offspring is created by μ times choosing an individual $s \in P_t$ uniformly at random and applying *standard bit mutation* on s, i.e. each bit is flipped independently with probability $1/n$. During the survival selection, the parent and offspring populations P_t and Q_t are merged into R_t. Then R_t is partitioned into layers $F_{t+1}^1, F_{t+1}^2, \dots$ using the *non-dominated sorting algorithm* (Deb et al. 2002) where F_{t+1}^1 consists of all non-dominated individuals, and F_{t+1}^i for $i > 1$ of individuals only dominated by those from $F_{t+1}^1, \dots, F_{t+1}^{i-1}$. Then the critical rank i^* with $\sum_{i=1}^{i^*-1} |F_t^i| < \mu$ and $\sum_{i=1}^{i^*} |F_t^i| \geq \mu$ is determined. All individuals with a lower rank than i^* are included in P_{t+1}, while the remaining individuals are selected from $F_t^{i^*}$ using Algorithm 2. Hereby, a normalized objective function f^n is computed and then each individual with rank at most i^* is associated with reference points. For the first, we use the normalization procedure from (Wietheger and Doerr 2023) which can be also used for maximization problems as shown in (Opris et al. 2024). We omit detailed explanations as they are not needed for our purposes. For an m-objective function $f \colon \{0,1\}^n \to \mathbb{N}_0^m$, the normalized fitness vector $f^n(x) := (f_1^n(x), \dots, f_m^n(x))$ of a search point x is computed as

$$f_j^n(x) = \frac{f_j(x) - y_j^{\min}}{y_j^{\text{nad}} - y_j^{\min}}$$

for each $j \in [m]$ where $y^{\text{nad}} := (y_1^{\text{nad}}, \dots, y_m^{\text{nad}})$ and $y^{\min} := (y_1^{\min}, \dots, y_m^{\min})$ from the objective space are called *nadir* and *ideal* points, respectively. Computing the nadir point is not trivial and we have $y_j^{\text{nad}} \geq \varepsilon_{\text{nad}}$, and $y_j^{\min} \leq y_j^{\text{nad}} \leq y_j^{\max}$ for every $j \in [m]$ where ε_{nad} is a positive threshold set by the user (see (Blank, Deb, and Roy 2019) or (Wietheger and Doerr 2023) for the details). Further, $y_j^{\max}$ and $y_j^{\min}$ are the maximum and minimum value in objective j from all search points seen so far (i.e. from $P_0, Q_0, \dots, P_t, Q_t$). After computing the normalisation, each individual x is associated with the reference point $\text{rp}(x)$ such that the distance between $f^n(x)$ and the line through the origin and $\text{rp}(x)$ is minimal. We use the same set of reference points $\mathcal{R}_p$ as proposed in (Deb and Jain 2014), originated in (Das and Dennis 1998). The points $\mathcal{R}_p$ are defined as

$$\left\{ \left(\frac{a_1}{p}, \dots, \frac{a_m}{p} \right) \;\middle|\; (a_1, \dots, a_m) \in \mathbb{N}_0^m, \sum_{i=1}^m a_i = p \right\}$$

where $p \in \mathbb{N}$ is a parameter one can choose according to the fitness function f. These are uniformly distributed on the simplex determined by the unit vectors $(1,0,\dots,0)^\intercal, (0,1,\dots,0)^\intercal, \dots, (0,0,\dots,1)^\intercal$. We chose this approach because the objective vectors of OMM are evenly distributed in the objective space, ensuring a well-spread set of solutions. Then, one iterates through all reference points choosing the one with the fewest already selected individuals for the next generation P_{t+1} is chosen. A reference point is omitted if it only has associated individuals already selected for P_{t+1}, and ties are broken uniformly at random. From the unselected individuals associated with the chosen reference point, the one closest to it is added to P_{t+1}, with ties again broken uniformly at random. Selection ends once the required number of individuals is reached (i.e. if $|Y_t| + |\tilde{F}_t^{i^*}| = \mu$).

Our analysis relies on the fact that NSGA-III preserves Pareto-optimal solutions: if the population exceeds the number of mutually incomparable solutions, and $p \geq 2m^{3/2} f_{\max}$, a covered vector on the Pareto-front remains covered in future iterations, as individuals with different objective vectors are associated to different reference points (see Lemma 2 in (Opris et al. 2024) for the details). The next lemma combines this observation with Lemma 3.4 from (Opris 2025a), which provides a general statement about the cover number of a vector on the Pareto front.

Lemma 1. *Consider NSGA-III on an m-objective function f with $\varepsilon_{nad} \geq f_{\max}$ and a set $\mathcal{R}_p$ of reference points for $p \in \mathbb{N}$ with $p \geq 2m^{3/2} f_{\max}$. Denote by P_t the current population and by F the Pareto front of f. Let S be a maximum set of mutually incomparable solutions and suppose that $\mu \geq |S|$. Then the following properties hold.*

(1) If $\mu \geq |S|$ then for each $x \in F_t^1$ there is an $y \in P_{t+1}$ weakly dominating x.
(2) Let $v \in F$ and $0 \leq \alpha \leq \mu/|S|$. If $c_t(v) \geq \alpha$ then also $c_{t+1}(v) \geq \alpha$.
(3) Let $v \in F$ and suppose that $c_{t+1}(v) < c_t(v)$. Then $c_{t+1}(w) \leq c_t(v)$ for every $w \in F$.
(4) Suppose that every $x \in P_t$ is Pareto-optimal. Then $d_t := \max\{c_t(v) \mid v \in F\}$ does not increase.

The m-OMM benchmark, originated in (Zheng and Doerr 2024a), is defined as follows.

Definition 2. *Let m be divisible by 2 and let the problem size be a multiple of $m/2$. Then the m-objective function m-OMM is defined by $m\text{-OMM} : \{0,1\}^n \to \mathbb{N}_0^m$ as*

$$m\text{-OMM}(x) = (f_1(x), \ldots, f_m(x))$$

with

$$f_\ell(x) = \begin{cases} \sum_{i=1}^{2n/m} x_{i+n(\ell-1)/m}, & \text{if } \ell \text{ is odd,} \\ \sum_{i=1}^{2n/m} (1 - x_{i+n(\ell-2)/m}), & \text{else,} \end{cases}$$

for all $x = (x_1, \ldots, x_n) \in \{0,1\}^n$.

In m-OMM, the bit string is divided into $m/2$ blocks, where f_{2j-1} and f_{2j} correspond to block j. Specifically, f_{2j-1} counts the number of ones, and f_{2j} counts the number of zeros in block j. Every search point is Pareto-optimal, as the total sum of objectives of any bit string is n. A Pareto-optimal set thus, which is also a maximum set of mutually incomparable solutions, has cardinality $(2n/m+1)^{m/2}$, since for each block $j \in [m/2]$ there are at most $2n/m + 1$ many fitness values (f_{2j-1}, f_{2j}).

Population Dynamics of NSGA-III on 2-OMM

Bounding the Maximum Cover Number: First, we establish a general upper bound on the time required to cover a subset of the Pareto front of a given cardinality with high probability. Then, we provide an additional bound on the time needed to evenly distribute solutions across that subset.

Lemma 3. *Consider NSGA-III on $f := 2\text{-OMM}$ under the same conditions as in Lemma 1. (i.e. $p \geq 4\sqrt{2}n$ since $f_{\max} = n$ and $\mu \geq n+1$). Then for a given natural $\alpha \leq 3n/8$ there is $\mathcal{A} \subset F$ with cardinality α which is covered in 64α generations with probability at least $1 - e^{-\Omega(\alpha)}$.*

Proof. By a classical Chernoff bound the probability is at least $1 - e^{-\Omega(n)}$ that there is an individual x initialized with $f_j(x) \in [3n/8, 5n/8]$. Suppose that this happens and fix such an individual x_0. Let $\mathcal{A} := \{v \in F \mid v_i \in [f_i(x_0) - \alpha/2, f_i(x_0) + \alpha/2] \text{ for all } i \in \{1,2\}\} \subset [3n/8 - \alpha/2, 5n/8 + \alpha/2]$. Then we see that $|\mathcal{A}| \geq \alpha$. Fix a covered $v \in \mathcal{A}$ and another uncovered $w \in \mathcal{A}$. We show with probability at least $1 - e^{-\Omega(\alpha)}$ the vector w is covered after 64α generations. Let $B_t := \{x \in P_t \mid f(x) \in \mathcal{A}\}$ and $d_t := \min_{x \in B_t} |f_1(x) - w_1|$. Note that $0 \leq d_t \leq \alpha$. Further, we see that w is covered if $d_t = 0$ since if $f_1(x) = w_1$ then also $f_2(x) = w_2$ due to $f_1(x) + f_2(x) = w_1 + w_2 = n$. By Lemma 1(1), d_t cannot increase, but it can be decreased in one single trial by choosing $x \in P_t$ with $|f_1(x) - w_1| = d_t$ as parent (probability at least $1/\mu$) and then flipping a one bit to zero and not changing any other bit if $f_1(x) - w_1 > 0$. On the other hand, if $f_1(x) - w_1 < 0$, flip a zero bit to one. Both happen with probability at least $(3n/8 - \alpha/2)/n \cdot (1 - 1/n)^{n-1} \geq (3n/8 - 3n/16)/(en) \geq 3/(16e)$. Then in one generation, d_t decreases with probability at least $1 - (1 - \frac{3}{16e\mu})^\mu \geq \frac{3/(16e)}{1+3/(16e)} \geq \frac{3}{32e} =: p$ where the first inequality is due to Lemma 10 in (Badkobeh, Lehre, and Sudholt 2015). Note that $p = \Omega(1)$ and for each $i \in [\alpha]$, define the random variable X_i as the number of generations such that $d_t = i$. Then the time until $d_t = 0$ is $X := \sum_{i=1}^\alpha X_i$, the latter stochastically dominated by the independent sum $Y := \sum_{i=1}^\alpha Y_i$ of geometrically distributed random variables with success probability $p_i = p = \Omega(1)$. Note that $\mathrm{E}[Y] = \alpha/p$ and we obtain by Theorem 1 in (Witt 2014) for $s := \sum_{i=1}^\alpha 1/p_i^2 = 1024\alpha e^2/9$ and $\lambda \geq 0$ the inequality $\Pr(Y \geq \mathrm{E}[Y] + \lambda) \leq \exp(-\frac{1}{4}\min\{\frac{\lambda^2}{s}, \lambda p\})$. For $\lambda = \alpha/p$ we obtain $\Pr(X \geq 64\alpha) \leq \Pr(X \geq 64e\alpha/3) = \Pr(X \geq 2\alpha/p) \leq \Pr(Y \geq 2\alpha/p) = \Pr(Y \geq \mathrm{E}[Y] + \alpha/p) \leq e^{-\Omega(\alpha)}$. By a union bound on all $w \in \mathcal{A}$ we see that $\mathcal{A}$ is covered in 64α generations with probability $1 - e^{-\Omega(\alpha)}$. □

Now we give the following upper bound on the time such that, with high probability, the solutions are evenly spread on a set $\mathcal{A}$ with cardinality α or, in other words, the cover number of each $v \in \mathcal{A}$ is bounded by $\lceil \mu/\alpha \rceil$ from above.

Lemma 4. *Consider NSGA-III on $f := 2\text{-OMM}$ and suppose that all conditions of Lemma 1 are satisfied. Suppose that $\mu = poly(n)$. Let $\alpha \leq 3n/8$ be a natural number and let $\gamma := \min\{\lceil n/\ln(n) \rceil, \lceil \mu/\alpha \rceil\}$. Then after $84\alpha + 46\gamma$ generations, each $v \in \mathbb{N}_0^m$ has cover number at most $\lceil \mu/\alpha \rceil$*

with probability $1-o(1)$*. Hence, if* $\alpha \geq n/\ln(n)$ *then* $\gamma \leq \alpha$ *and therefore,* 130α *generations suffice. The expected number of generations is* $O(\alpha+\gamma)$*.*

Proof. By Lemma 3 there is a set $\mathcal{A} \subset P$ with cardinality α which is covered after 64α generations with probability at least $1-e^{-\Omega(\alpha)}$. Suppose that this happens. Denote the decrease of the cover number $c_t(v)$ of a vector $v \in F$ as a *success*, if $c_t(v) \leq \lceil \mu/\alpha \rceil$. If a success occurs, we see that $c_{t+1}(w) \leq c_t(v) \leq \lceil \mu/\alpha \rceil$ of all other Pareto-optimal vectors w by Lemma 1(3) and it cannot increase by Lemma 1(4) (since every solution is Pareto-optimal), implying the statement of the lemma. We show with probability $1-o(1)$ that all $v \in \mathcal{A}$ have a cover number of at least $\lceil \mu/\alpha \rceil$ or a success occurred after further $46\gamma+20\alpha$ generations. In the former case $\lceil \mu/\alpha \rceil = \mu/\alpha$, all other $v \in F \setminus \mathcal{A}$ have cover number 0, and hence, the cover number of all $v \in F$ is also bounded by $\lceil \mu/\alpha \rceil$. Depending on the value of γ, we consider two cases where F denotes the Pareto front of 2-OMM.

Case 1: Let $\lceil \mu/\alpha \rceil \leq \lceil n/\ln(n) \rceil$ (i.e. $\gamma = \lceil \mu/\alpha \rceil$). Fix $v \in F$, denote by c_t its cover number and for $j \in [\gamma-1]$ let X_j be a random variable that counts the number of generations with $c_t = j$. Then the number of generations until a success occurs or the cover number of v is at least γ is at most $X := \sum_{j=1}^{\gamma-1} X_j$. Note that c_t can be increased by choosing an individual y with $f(x) = f(y)$ as parent and flipping no bits (prob. $1/\mu \cdot (1-1/n)^n \geq 1/(4\mu) =: \sigma_t$). Hence, the probability of increasing c_t in one generation is at least $1-(1-\sigma_t)^\mu \geq \frac{\sigma_t \mu}{1+\sigma_t\mu} = \frac{1/4}{1+1/4} = \frac{1}{5}$. Hence, X is stochastically dominated by an independent sum $Z := \sum_{j=1}^{\gamma-1} Z_j$ of geometrically distributed random variables Z_j with parameter $p = 1/5$. Then $\mathrm{E}[X] \leq \mathrm{E}[Z] \leq 5\gamma$ and hence, by Theorem 1 in (Witt 2014), we obtain for $s := \sum_{i=1}^{\gamma-1} 1/p_i^2 \leq 25\gamma$, and $\lambda \geq 0$ the inequality $\Pr(Z \geq \mathrm{E}[Z]+\lambda) \leq \exp(-\frac{1}{4}\min\{\frac{\lambda^2}{s}, \lambda p\})$ and for $\lambda = 40\gamma+20\alpha$ we obtain $\Pr(X \geq (5+40)\gamma+20\alpha) = \Pr(X \geq 5\gamma+(40\gamma+20\alpha)) \leq \Pr(Z \geq \mathrm{E}[Z]+40\gamma+20\alpha) \leq e^{-2\gamma-\alpha}$. By a union bound on $|\mathcal{A}| = \alpha$ different Pareto-optimal vectors, we see that with probability at most $\alpha \cdot e^{-2\gamma-\alpha} = o(1)$ a success occurred or the cover number of all $v \in \mathcal{A}$ is at least $\lceil \mu/\alpha \rceil$ after further $45\gamma+20\alpha$ generations.

Case 2: Suppose that $\lceil \mu/\alpha \rceil > \lceil n/\ln(n) \rceil$ (i.e. $\gamma = \lceil n/\ln(n) \rceil$). Fix $v \in F$, and let Y_t denote the number of individuals x such that $f(x) = v$ at generation t. Then with probability $1-e^{-2\gamma-\alpha}$ after further $45\gamma+20\alpha$ generations by Case 1, a success occurred or the cover number of v is at least $n\ln(n)$. Suppose the latter (otherwise the statement of the lemma holds) and let Z_t be the number of newly created individuals with fitness vector v in generation t. We have $\mathrm{E}[Z_t] \geq Y_t/4 \geq n/(4\ln(n))$: A generation consists of μ independent trials and in each trial, with probability at least $n/(\ln(n)\mu)$, an individual x with $f(x) = v$ is selected as the parent, and during mutation, no bit is flipped with probability at least $(1-1/n)^n \geq 1/4$. Hence, by a classical Chernoff bound, $\Pr(Z_t \leq 3/5 \cdot \mathrm{E}[Z_t]) = \Pr(Z_t \leq (1-2/5)\cdot \mathrm{E}[Z_t]) \leq e^{-\Omega(\mathrm{E}[Z_t])} = e^{-\Omega(n/\ln(n))}$. Hence, with probability at least $1-e^{-\Omega(n/\ln(n))}$, we have $Y_{t+1} \geq \min\{Y_t + 3/5 \cdot \mathrm{E}[Z_t], \lceil \mu/\alpha \rceil\} = \min\{Y_t + 3Y_t/20, \lceil \mu/\alpha \rceil\} = \min\{23Y_t/20, \lceil \mu/\alpha \rceil\}$. In other words, Y_t increases by a factor of at least $23/20$, unless the value $\lceil \mu/\alpha \rceil$ has already been reached. Note that for n large enough $\lceil n/\ln(n) \rceil$ such generations in succession are sufficient to reach a cover number of v of at least $\lceil \mu/\alpha \rceil$ or a success occured, since $n/\ln(n) \cdot (23/20)^{n/\ln(n)} = \omega(\mu)$. Moreover, such a sequence of generations occurs with probability at least $1-e^{-\Omega(n/\ln(n))}$ by a union bound on all these generations. By a further union bound on all $v \in \mathcal{A}$, we see that after $\lceil n/\ln(n) \rceil = \gamma$ generations, a success occurred or the cover number of each $v \in F$ is at least $\lceil \mu/\alpha \rceil$ with probability $1-o(1)$. This also leading to the desired $46\gamma+20\alpha$ generations in this case.

Hence, in any case we see that with probability $1-o(1)$, after $\kappa := 64\alpha+46\gamma+20\alpha = 84\alpha+46\gamma$ generations, each $v \in F$ has cover number at most $\lceil \mu/\alpha \rceil$ with probability $1-o(1)$. If this does not happen, we repeat the arguments from either Case 1 or Case 2 for another period of κ generations, including the preceding phase from Lemma 3 to cover $\mathcal{A}$ if necessary. The expected number of periods is $1+o(1)$, concluding the proof. □

Controlling the Exploration of Search Points: First, we bound the spread of solutions in $O(n/\ln(n))$ generations.

Lemma 5. *Consider NSGA-III on* 2*-OMM under the same conditions as in Lemma 1. Suppose that* $\mu = poly(n)$ *and let* $c > 0$ *be a constant. Then, after* $cn/\ln(n)$ *generations, there is no* $y \in P_t$ *with* $|y|_1 \geq 3n/4$ *with probability* $1-o(1)$*.*

Proof. Let $d_t := \min\{\max\{3n/4 - |y|_1, 0\} \mid y \in P_t\}$. By a classical Chernoff bound each individual x satisfies $3n/8 < |x|_1 < 5n/8$ with probability $1-\mu e^{-\Omega(n)} = 1-o(1)$ after initialization. Suppose that this happens. Then $d_0 \geq n/8$ and therefore, in order to create an individual y with $|y|_1 \geq 3n/4$ within $\lceil cn/\ln(n) \rceil \leq 2cn/\ln(n)$ generations, it is necessary that d_t reaches 0, particularly decreases by at least $\ln(n)/(16c)$ in one such iteration. This requires that at least $\ell := \lceil \ln(n)/(16c) \rceil$ many zero bits are flipped simultaneously in one individual. The latter happens with probability at most $\binom{n}{\ell}\left(\frac{1}{n}\right)^\ell = \frac{n!}{\ell!(n-\ell)!n^\ell} \leq \frac{1}{\ell!} \leq \frac{e^\ell}{\ell^\ell} = e^{-\omega(\ell)} = e^{-\omega(\ln(n))}$ in one single trial where the last inequality is due to Stirling's formula. By a union bound on at most $\mu\lceil cn/\ln(n) \rceil$ mutation steps after $cn/\ln(n)$ generations, we see that the probability is $o(1)$ to decrease d_t by at least ℓ one time within $\lceil cn/\ln(n) \rceil$ generations (since $\mu = \mathrm{poly}(n)$), concluding the proof. □

Second, we bound the time required for the maximum number of ones among individuals in P_t to increase from $n-n^b$ to $n-n^a$, for constants $0 \leq a < b \leq 3/4$.

Lemma 6. *Consider NSGA-III on* 2*-OMM under the same conditions as in Lemma 1. Let* $0 \leq a < b \leq 3/4$ *be two constants. Assume that the maximum cover number is at most* $\beta = o(n^{1-b})$*. Suppose every* $x \in P_t$ *satisfies* $|x|_1 \leq n - n^b$*. Then with probability* $1-o(1)$ *NSGA-III requires more than* $(b-a)n\ln(n)/(32e\beta)$ *generations to create an individual* x *with* $|x|_1 \geq n-n^a$*. Hence, the expected number of generations is at least* $\Omega(n\ln(n)/\beta)$*.*

Proof. Consider $Y_t := \max\{\max\{|x|_1, n - \lceil n^b \rceil\} \mid x \in P_t\}$. If all individuals x satisfy $|x|_1 \le n - \lceil n^b \rceil$ (which is the case at the beginning) then $Y_t = n - \lceil n^b \rceil$. We created an x with $|x|_1 \ge n - n^a$ if $Y_t \ge n - \lfloor n^a \rfloor$. At first we bound the probability p^* to increase Y_t by at least 8 in one generation from above as follows. In one single trial, for each $i \in \{0, \ldots, Y_t\}$, one can choose an $x \in P_t$ with $|x|_1 = Y_t - i$ (i.e. $|x|_0 = n - Y_t + i$) (prob. at most β/μ) and then flip $i + 8$ zero bits (prob. at most $\binom{n-Y_t+i}{i+8} \cdot 1/n^{i+8}$). By a union bound on all i, we obtain that Y_t increases by at least 8 in a single trial with probability at most $\frac{\beta}{\mu} \sum_{i=0}^{Y_t} \binom{n-Y_t+i}{i+8} \frac{1}{n^{i+8}} \le \frac{\beta}{\mu} \sum_{i=0}^{Y_t} \binom{\lceil n^b \rceil + i}{i+8} \frac{1}{n^{i+8}} \le \frac{\beta}{\mu} \left(\frac{\lceil n^b \rceil}{n} \right)^8 \sum_{i=0}^{Y_t} \frac{(\lceil n^b \rceil + i) \cdot \ldots \cdot (\lceil n^b \rceil + 1)}{n^i (i+8)!} \le \frac{\beta}{\mu} \left(\frac{\lceil n^b \rceil}{n} \right)^8$ (where we used $\lceil n^b \rceil + i \le 2n$ and $\sum_{i=0}^{Y_t} \frac{2^i}{(i+8)!} \le 1$). Hence, by a union bound on μ single trials, we obtain the inequality $p^* \le \beta \cdot (\lceil n^b \rceil / n)^8 \le 256\beta/n^2$ (since $b \le 3/4$ as well as $\lceil r \rceil \le 2r$ for all $r \ge 1$ is satisfied). Again by a union bound, Y_t changed by at least 8 after $(b - a)n \ln(n)/(32a\beta)$ generations with probability $o(1)$. So we assume that Y_t is never changed by at least 8 and for $Y_t \in [n - \lceil n^b \rceil, \ldots, n - \lfloor n^a \rfloor]$ and natural $1 \le \ell \le (\lceil n^b \rceil - \lfloor n^a \rfloor)/8$ let X_ℓ be the random variable which counts the number of generations with $Y_t \in \{n - \lceil n^b \rceil + 8(\ell - 1), \ldots, n - \lceil n^b \rceil + 8\ell - 1\}$. Now, for $k := k(n, \ell) := \lceil n^b \rceil - 8(\ell - 1)$ we justify that X_ℓ stochastically dominates a geometrically distributed random variable Z_ℓ with success probability $p_\ell = 1.5e\beta k/n$:

A necessary condition that Y_t leaves $\{n - k, \ldots, n - \lceil n^b \rceil + 8\ell - 1\}$ is that Y_t increases by one in a generation which happens with probability at most $\frac{\beta}{\mu} \sum_{i=0}^{Y_t} \binom{n-Y_t+i}{i+1} \frac{1}{n^{i+1}} \le \frac{\beta}{\mu} \sum_{i=0}^{Y_t} \binom{k+i}{i+1} \frac{1}{n^{i+1}} \le \frac{\beta}{\mu} \frac{k}{n} \sum_{i=0}^{Y_t} \frac{(k+i) \cdot \ldots \cdot (k+1)}{n^i (i+1)!} \le \frac{1.5e\beta k}{\mu n}$ (by choosing a parent x with $|x|_1 = Y_t - i$ and then flipping $i + 1$ zero bits for $i \in \{0, \ldots, Y_t\}$). For the last inequality we used $k + i \le (1 + \ln(1.5))n$ for all $i \in \{0, \ldots, Y_t\}$ and n sufficiently large, and $\sum_{i=0}^{Y_t} \frac{(1+\ln(1.5))^i}{(i+1)!} \le \sum_{i=0}^{\infty} \frac{(1+\ln(1.5))^i}{i!} = e^{1+\ln(1.5)} = 1.5e$. Then apply a union bound on μ trials to finish the justification.

This implies that for $\delta := \delta(a, b, n) := \lfloor \frac{\lceil n^b \rceil - \lfloor n^a \rfloor}{8} \rfloor$ the number T of generations until $Y_t \ge n - \lfloor n^a \rfloor$ (which is at least $\sum_{\ell=1}^{\delta} X_\ell$) stochastically dominates the independent sum $Z := \sum_{\ell=1}^{\delta} Z_\ell$ of geometrically distributed random variables Z_ℓ. Note also that $\ln(n) \le \sum_{i=1}^{n} 1/i \le \ln(n) + 1$ and therefore $\sum_{i=1}^{n} 1/i - \sum_{i=1}^{q} 1/i \ge \ln(n) - (\ln(q) + 1) = \ln(n/q) - 1$ for $q \in [n]$. Therefore, for $\kappa := \lceil \lceil n^b \rceil / 8 \rceil$ we obtain $\mathrm{E}[Z] = \sum_{\ell=1}^{\delta} \mathrm{E}[Z_\ell] = \sum_{\ell=1}^{\delta} \frac{1}{p_\ell} = \sum_{\ell=1}^{\delta} \frac{n}{1.5e\beta(\lceil n^b \rceil - 8(\ell-1))} = \frac{n}{12e\beta} \sum_{\ell=0}^{\delta-1} \frac{1}{\lceil n^b \rceil / 8 - \ell} \ge \frac{n}{12e\beta} \sum_{\ell=0}^{\delta-1} \frac{1}{\kappa - \ell} \ge \frac{n}{12e\beta} \left(\sum_{\ell=0}^{\kappa-1} \frac{1}{\kappa-\ell} - \sum_{\ell=\delta}^{\kappa-1} \frac{1}{\kappa-\ell} \right) \ge \frac{n}{12e\beta} \left(\sum_{\ell=1}^{\kappa} \frac{1}{\ell} - \sum_{\ell=1}^{\kappa-\delta} \frac{1}{\ell} \right) \ge \frac{n}{12e\beta} \left(\ln\left(\frac{\kappa}{\kappa-\delta} \right) - 1 \right) \ge \frac{n}{12e\beta} \left(\ln\left(\frac{\lceil n^b \rceil/8+1}{\lfloor n^a \rfloor/8+2} \right) - 1 \right) \ge \frac{n}{12e\beta} \left(\ln\left(\frac{n^b+8}{n^a+16} \right) - 1 \right) \ge \frac{(b-a)n\ln(n)}{16e\beta}$ for n sufficiently large. Then, under the condition that Y_t never changes by at least 8 within $(b - a)n \ln(n)/(32e\beta)$ generations, we see by Theorem 1 in (Witt 2014) that for $\lambda := \mathrm{E}[Z]/2$ and $s := \sum_{\ell=1}^{\delta} 1/p_\ell^2 = \sum_{\ell=1}^{\delta} \frac{n^2}{2.25e^2\beta^2(\lceil n^b \rceil - 8(\ell-1))^2} \le \sum_{j=1}^{\infty} \frac{n^2}{2.25e^2\beta^2 j^2} \le \frac{n^2\pi^2}{13.5e^2\beta^2}$ (due to $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$) the inequality $\Pr(T \le \frac{(b-a)n\ln(n)}{32e\beta}) \le \Pr(Z \le \mathrm{E}[Z]/2) = \Pr(Z \le \mathrm{E}[Z] - \mathrm{E}[Z]/2) \le \exp(-\lambda^2/(2s)) = o(1)$ holds. This proves the lemma with the law of total probability. □

A Lower Runtime Bound

In this section, we establish the desired lower bound on the runtime of NSGA-III on the 2-OMM problem by putting together the results from the previous section. The core of the lower bound argument relies on an iterative process that repeatedly applies Lemmas 4 and 6 to reduce the maximum cover number β as the population approaches the Pareto front's extremes.

Theorem 7. *Consider NSGA-III on 2-OMM under the same conditions as in Lemma 1. Further suppose that $n+1 \le \mu \in O(\ln(n)^c n)$ for a constant $0 < c < 1$. Then the expected number of generations to cover the whole Pareto front is at least $\Omega(n^2 \ln(n)/\mu)$.*

Proof. Fix a constant $\chi > 0$ such that $\mu \le \chi \ln(n)^c n$ for n sufficiently large. At first we see by Lemma 5 that with probability $1 - o(1)$ there is no individual y in P_t with $|y|_1 \ge 3n/4$ within $130\lfloor n/\ln(n) \rfloor$ generations. Further, by Lemma 4 on $\alpha = \lfloor n/\ln(n) \rfloor$, we obtain that, after 130α generations, the maximum cover number is at most $\lceil \mu/\alpha \rceil \le \frac{\mu}{n/\ln(n)-1} + 1 \le \frac{2\mu\ln(n)}{n-\ln(n)} \le 4\chi \ln(n)^{1+c}$ with probability $1 - o(1)$. Suppose that this happens. We now apply Lemma 6 with $b = 3/4$ and $a = 1/2$ to obtain with probability $1 - o(1)$, that after further $(b - a)n \ln(n)/(32e \cdot 4\chi \ln(n)^{1+c}) = n/(512\chi e \ln(n)^c) = d_0 n/\ln(n)^c$ generations for $d_0 = 1/(512\chi e)$, no solution x with $|x|_1 \ge n - n^{1/2}$ is created. Also apply Lemma 4 on that number of generations for $\alpha = \lfloor d_0 n/(130 \ln(n)^c) \rfloor$ to obtain for $e_0 := 260\chi/d_0$ that with probability $1 - o(1)$ the maximum cover number is at most $\lceil \mu/\alpha \rceil \le \lceil \chi \ln(n)^c n/\alpha \rceil \le e_0 \ln(n)^{2c} = \max\{e_0 \ln(n)^{2c}, 16\mu/(3n)\}$ for n sufficiently large (the latter equality holds due to $e_0 \ln(n)^{2c} \in \omega(\mu/n)$).

Suppose that these two happen. In the following, we iteratively reduce the maximum cover number as the population approaches the extreme solution 1^n. To this end, let $\ell := \lceil (2c + 1)/(1 - c) \rceil \in O(1)$ and suppose that for $j \in \{0, \ldots, \ell - 1\}$ there are constants $0 < b_j \le 1/2$ and $d_j, e_j \ge 0$ such that after $d_j n \ln(n)^j / \ln(n)^{(2+j-1)c}$ generations, no solution x with $|x|_1 \ge n - n^{b_j}$ is created, that $(\ln(n))^{(2+j)c-j} = \omega(\mu/n)$ and the maximum cover number is at most $\beta = e_j (\ln(n))^{(2+j)c-j} = \max\{e_j (\ln(n))^{(2+j)c-j}, 16\mu/(3n)\}$ (where the case $j = 0$ already occurred). Now fix a further constant b_{j+1} with $1/8 < b_{j+1} < b_j$. Then again by Lemma 6 we see that with probability $1 - o(1)$ in $\frac{(b_j - b_{j+1})n\ln(n)}{32e \cdot \beta} = \frac{(b_j - b_{j+1})n\ln(n)}{32e \cdot e_j (\ln(n))^{(2+j)c-j}} = \frac{d_{j+1} n \ln(n)^{j+1}}{\ln(n)^{(2+j)c}}$ generations for $d_{j+1} = \frac{b_j - b_{j+1}}{32e \cdot e_j}$ no solution x with $|x|_1 \ge$

$n - n^{b_{j+1}}$ is created. After this time, by Lemma 4 on $\alpha = \min\{\lfloor \frac{d_{j+1} n \ln(n)^{j+1}}{(130 \ln(n)^{(2+j)c})} \rfloor, \lfloor 3n/8 \rfloor\}$, the maximum cover number is at most $\lceil \mu/\alpha \rceil \le \max\{\frac{260\mu \ln(n)^{(2+j)c}}{d_{j+1} n \ln(n)^{j+1}}, \frac{16\mu}{3n}\} \le \max\{\frac{260\chi n \ln(n)^{(3+j)c}}{d_{j+1} n \ln(n)^{j+1}}, \frac{16\mu}{3n}\} = \max\{\frac{e_{j+1} \ln(n)^{(3+j)c}}{\ln(n)^{j+1}}, \frac{16\mu}{3n}\}$ for $e_{j+1} := 260\chi/d_{j+1}$ with probability $1 - o(1)$. If $\ln(n)^{(3+j)c-(j+1)} = \omega(\mu/n)$, we increase j by one and repeat this argument. We stop when $\ln(n)^{(3+j)c-(j+1)} = O(\mu/n)$. Since $(3+\ell)c - (\ell+1) \le 0$, we have at most $\ell = O(1)$ such repetitions. After the last repetition we have that $\alpha = \Omega(n)$. Hence, by applying a union bound on all repetitions, we conclude that with probability $1 - o(1)$, there exists a generation t^{spread} such that no individual x with $|x|_1 \le n - n^{1/8}$ is created and the maximum cover number is at most $e_{\text{spread}}\mu/n$ for a constant $e_{\text{spread}} > 0$. Suppose this event occurs, and apply Lemma 6 once more with $b = 1/8$ and $a = 1/16$. This yields that, after $\Omega(n \ln(n)/(e_{\text{spread}}\mu/n)) = \Omega(n^2 \ln(n)/\mu)$ generations in expectation (from time t^{spread} onward), a search point x with $|x|_1 \ge n - n^{1/16}$ is created, concluding the proof. □

An Improved Upper Runtime Bound

To complement our analysis, we establish an improved upper bound on the expected runtime of NSGA-III on m-OMM for a constant number of objectives m. Our approach closely follows the methodology provided by (Opris et al. 2024), with the added consideration of the cover number.

Theorem 8. *Consider NSGA-III on m-OMM for a constant number m of objectives under the same conditions as in Lemma 1 with population size $S_m := (2n/m+1)^{m/2} \le \mu \le \ln(n)S_m$. Then a Pareto-optimal set of m-OMM is found in expected $O(S_m n \ln n/\mu + n\mu/S_m)$ generations or, in other words, in expected $O(S_m n \ln n + n\mu^2/S_m)$ fitness evaluations. If $\mu \ge \ln(n)S_m$, a Pareto set is found in expected $O(n \ln(n))$ generations (which is Theorem 5.2 in (Opris et al. 2024)).*

Proof. Suppose that $S_m \le \mu \le \ln(n)S_m$. Fix a vector v on the Pareto front. We estimate the probability not to cover v after $6S_m en \ln(n)/\mu + 10n\lceil \mu/S_m \rceil$ generations. For each generation t let $d_t := \min_{x \in P_t} \sum_{j=1}^{m/2} |f_{2j-1}(x) - v_{2j-1}|$. Note that $0 \le d_t \le n$ and that we have covered v if $d_t = 0$. Further, by Lemma 1(1), d_t cannot increase. Let $y \in P_t$ be with $\sum_{j=1}^{m/2} |f_{2j-1}(y) - v_{2j-1}| = d_t$. We first increase the cover number of $f(y)$ to $\lfloor \mu/S_m \rfloor$ (which, by Lemma 1(2), can only decrease if it exceeds $\lfloor \mu/S_m \rfloor$, and even then not below this value), and then proceed to decrease d_t. The latter then happens with probability at least $\lfloor \mu/S_m \rfloor/\mu \cdot i/n \cdot (1-1/n)^{n-1} \ge i/(2S_m en)$ in one single trial and hence, with probability at least $1 - (1 - i/(2S_m en))^\mu \ge \frac{i\mu/(2S_m en)}{i\mu/(2S_m en)+1} =: p_i$ in one generation. Hence, the time T until $d_t = 0$ is stochastically dominated by an independent sum of geometrically distributed random variables Y_i , Z_j ($i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \nu\}$ for $\nu := n \cdot \lfloor \mu/S_m \rfloor$) with success probability p_i and $\tilde{p} = 1/5$ respectively (compare also with the proof of Lemma 4 for the latter). Let $Y := \sum_{i=1}^n Y_i$ and $Z := \sum_{j=1}^\nu Z_j$. We have $\mathrm{E}[Y] = \sum_{i=1}^n 1/p_i = \sum_{i=1}^n (1 + S_m en/(i\mu)) \le n + S_m en(\ln(n)+1)/\mu \le 2S_m en \ln(n)/\mu$ for n sufficiently large and $\mathrm{E}[Z] = 5n \cdot \lfloor \mu/S_m \rfloor$. By Theorem 1 in (Witt 2014) we obtain for $s = \sum_{i=1}^n 1/p_i^2 = \sum_{i=1}^n (1 + 2S_m en/(i\mu))^2$, $p = \min_{i \in [n]} p_i = p_1 \ge \mu/(4S_m en)$, and $\lambda = 8mS_m en \ln(n)/\mu$ that $\Pr(Y \ge \mathrm{E}[Y] + \lambda) \le \exp\left(-\frac{1}{4}\min\left\{\frac{\lambda^2}{s}, \lambda p\right\}\right) \le n^{-2m}$ since $\lambda^2/s = \Omega(\ln^2(n))$ and $\lambda p \ge 2m \ln(n)$. Further, we see for $\tilde{s} = 25\nu$, and $\tilde{\lambda} = \mathrm{E}[Z]$ that $\Pr(Z \ge \mathrm{E}[Z] + \tilde{\lambda}) \le \exp(-\frac{1}{4}\min\{\frac{\tilde{\lambda}^2}{\tilde{s}}, \tilde{\lambda}\tilde{p}\}) = e^{-\Omega(n)}$. These two inequalities imply $\Pr(T \ge \mathrm{E}[Y] + \mathrm{E}[Z] + \lambda + \tilde{\lambda}) \le \Pr(Y + Z \ge \mathrm{E}[Y] + \mathrm{E}[Z] + \lambda + \tilde{\lambda}) \le \Pr(Y \ge \mathrm{E}[Y] + \lambda) + \Pr(Z \ge \mathrm{E}[Z] + \tilde{\lambda}) \le n^{-2m} + e^{-\Omega(n)}$. By a union bound on all possible v, the probability that there is a fitness vector v such that P_t does not contain a Pareto-optimal solution x with $f(x) = v$ after $\mathrm{E}[Y] + \mathrm{E}[Z] + \lambda + \tilde{\lambda} \le 10S_m en \ln(n)/\mu + 10n\lfloor \mu/S_m \rfloor = O(S_m n \ln(n)/\mu + \mu n/S_m)$ generations is at most $(2n/m+1)^{m/2} \cdot (n^{-2m} + e^{-\Omega(n)}) = o(1)$. If this does not happen we repeat all the above arguments. Note that in expectation, $1 + o(1)$ such periods are sufficient. □

This result improves the corresponding upper bound from (Opris et al. 2024) by a factor of $\min\{S_m/\mu, \mu/(S_m \ln(n))\}$ both in terms of generations and fitness evaluations if $S_m \le \mu = O(S_m \ln(n))$. Along with Theorem 7, we see in the case $m = 2$ that for $n + 1 \le \mu \le (n+1)\ln(n)^c$ for a constant $0 \le c \le 1/2$ the full Pareto front is covered in expected $\Theta(n^2 \ln(n)/\mu)$ generations, which is a tight runtime bound.

Conclusions

In this paper, we analyzed the widely used NSGA-III algorithm on the simple m-OMM problem and established lower runtime bounds for $m = 2$, as well as improved upper runtime bounds for a constant number m of objectives for a carefully chosen population size μ. For $m = 2$, this leads to a tight runtime bound, where NSGA-III outperforms NSGA-II, distributing solutions more evenly across the Pareto front. This is very surprising, since the latter is the state of the art algorithm for bi-objective problems (with around 60000 citations). Unlike previous work (Opris 2025a), where NSGA-III's dynamics were analyzed on m-OJZJ by first exploring the local optima, and then spreading the solutions evenly on the Pareto front, our analysis focused on the spread *during exploration*. Future research directions may include bounding the maximum cover number on benchmark problems, where it is necessary to *reach* the Pareto front at first glance, like the classic COUNTINGONESCOUNTINGZEROS (COCZ), or practical scheduling and graph problems. We hope that these insights provide a deeper understanding of the strengths and limitations of NSGA-III and may serve as a foundation for analyzing its behavior on more complex fitness landscapes. Ultimately, this understanding can aid practitioners in developing enhanced versions of the algorithm with improved performance for efficiently optimizing problems with complex and rugged fitness landscapes.

References

Amamou, Y.; and Jebari, K. 2023. Multiobjective Evolutionary Algorithms for Engineering Design Problems. In *International Conference on Advanced Intelligent Systems for Sustainable Development*, 318–331. Springer.

Andersson, J. 2003. Applications of a Multi-objective Genetic Algorithm to Engineering Design Problems. In *Evolutionary Multi-Criterion Optimization*, 737–751. Springer.

Badkobeh, G.; Lehre, P. K.; and Sudholt, D. 2015. Black-box Complexity of Parallel Search with Distributed Populations. In *Proceedings of the Foundations of Genetic Algorithms*, FOGA 2015, 3–15. ACM Press.

Bian, C.; Zhou, Y.; Li, M.; and Qian, C. 2023. Stochastic population update can provably be helpful in multi-objective evolutionary algorithms. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI 2023, 5513–5521. ijcai.org.

Blank, J.; Deb, K.; and Roy, P. C. 2019. Investigating the Normalization Procedure of NSGA-III. In *Evolutionary Multi-Criterion Optimization*, 229–240. Springer.

Cerf, S.; Doerr, B.; Hebras, B.; Kahane, Y.; and Wietheger, S. 2023. The First Proven Performance Guarantees for the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) on a Combinatorial Optimization Problem. In *Proceedings of the International Joint Conference on Artificial Intelligence*, IJCAI 2023, 5522–5530. ijcai.org.

Chaudhari, P.; Thakur, A. K.; Kumar, R.; Banerjee, N.; and Kumar, A. 2022. Comparison of NSGA-III with NSGA-II for Multi Objective Optimization of Adiabatic Styrene Reactor. *Materials Today: Proceedings*, 57: 1509–1514.

Dang, D.-C.; Opris, A.; Salehi, B.; and Sudholt, D. 2023. Analysing the Robustness of NSGA-II under Noise. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2023, 642–651. ACM Press.

Dang, D.-C.; Opris, A.; and Sudholt, D. 2024a. Crossover can guarantee exponential speed-ups in evolutionary multi-objective optimisation. *Artificial Intelligence*, 330: 104098.

Dang, D.-C.; Opris, A.; and Sudholt, D. 2024b. Level-Based Theorems for Runtime Analysis of Multi-objective Evolutionary Algorithms. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVIII, 246–263. Springer.

Dang, D.-C.; Opris, A.; and Sudholt, D. 2025a. Runtime Analysis of Functions Where Widely-Used Evolutionary Multi-Objective Algorithms Beat Simple Ones. *ACM Transactions on Evolutionary Learning and Optimization*.

Dang, D.-C.; Opris, A.; and Sudholt, D. 2025b. Why Dominance Is Not Enough: Lessons from Practical Evolutionary Multi-Objective Algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2025, 1604–1612. ACM.

Das, I.; and Dennis, J. E. 1998. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM Journal on Optimization*, 8(3): 631–657.

Deb, K.; and Jain, H. 2014. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Transactions on Evolutionary Computation*, 18(4): 577–601.

Deb, K.; Pratap, A.; Agarwal, S.; and Meyarivan, T. 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2): 182–197.

Deng, R.; Zheng, W.; Li, M.; Liu, J.; and Doerr, B. 2024. Runtime Analysis for State-of-the-Art Multi-objective Evolutionary Algorithms on the Subset Selection Problem. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVIII, 264–279. Springer.

Doerr, B.; Ivan, T.; and Krejca, M. S. 2025. Speeding Up the NSGA-II With a Simple Tie-Breaking Rule. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2025, 26964–26972. AAAI Press.

Doerr, B.; Krejca, M. S.; and Opris, A. 2025. Tight Runtime Guarantees From Understanding the Population Dynamics of the GSEMO Multi-Objective Evolutionary Algorithm. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI 2025, 8876–8884. ijcai.org.

Doerr, B.; and Qu, Z. 2022. A First Runtime Analysis of the NSGA-II on a Multimodal Problem. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVII, 399–412. Springer.

Doerr, B.; and Qu, Z. 2023a. From Understanding the Population Dynamics of the NSGA-II to the First Proven Lower Bounds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2023, 12408–12416. AAAI Press.

Doerr, B.; and Qu, Z. 2023b. Runtime Analysis for the NSGA-II: Provable Speed-ups from Crossover. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2023, 12399–12407. AAAI Press.

García, J. L. L.; Monroy, R.; Sosa Hernández, V. A.; and Coello Coello, C. A. 2021. COARSE-EMOA: An indicator-based evolutionary algorithm for solving equality constrained multi-objective optimization problems. *Swarm and Evolutionary Computation*, 67: 100983.

Handl, J.; Kell, D.; and Knowles, J. 2007. Multiobjective optimization in bioinformatics and computational biology. *IEEE/ACM Trans Comput Biol Bioinform*, 79: 279–292.

Khare, V.; Yao, X.; and Deb, K. 2003. Performance Scaling of Multi-objective Evolutionary Algorithms. In *Evolutionary Multi-Criterion Optimization*, 376–390. Springer.

Koziel, S.; Pietrenko-Dabrowska, A.; and Szczepanski, S. 2025. Multi-objective artificial-intelligence-based parameter tuning of antennas using variable-fidelity machine learning. *Scientific Reports*, 15: 21693.

Luukkonen, S.; van den Maagdenberg, H. W.; Emmerich, M. T.; and van Westen, G. J. 2023. Artificial Intelligence in Multi-Objective Drug Design. *Current Opinion in Structural Biology*, 79: 102537.

Mkaouer, W.; Kessentini, M.; Shaout, A.; Koligheu, P.; Bechikh, S.; Deb, K.; and Ouni, A. 2015. Many-Objective Software Remodularization Using NSGA-III. *ACM Transactions on Software Engineering and Methodology*, 24(3).

Mukhopadhyay, A.; Ray, S.; Maulik, U.; and Bandyopadhyay, S. 2024. *Multiobjective Optimization Algorithms for Bioinformatics*. Springer Singapore.

Opris, A. 2025a. A First Runtime Analysis of NSGA-III on a Many-Objective Multimodal Problem: Provable Exponential Speedup via Stochastic Population Update. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI 2025, 8903–8911. ijcai.org.

Opris, A. 2025b. A First Runtime Analysis of the PAES-25: An Enhanced Variant of the Pareto Archived Evolution Strategy. In *Proceedings of the 18th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*, FOGA 2025, 202–213. ACM.

Opris, A. 2025c. A Many Objective Problem Where Crossover is Provably Indispensable. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2025, 27108–27116. AAAI Press.

Opris, A.; Dang, D.-C.; Neumann, F.; and Sudholt, D. 2024. Runtime Analyses of NSGA-III on Many-Objective Problems. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2024, 1596–1604. ACM Press.

Qu, Q.; Ma, Z.; Clausen, A.; and Jørgensen, B. N. 2021. A Comprehensive Review of Machine Learning in Multi-objective Optimization. In *2021 IEEE 4th International Conference on Big Data and Artificial Intelligence*, BDAI 2021, 7–14.

Schneider, L.; Bischl, B.; and Thomas, J. 2023. Multi-Objective Optimization of Performance and Interpretability of Tabular Supervised Machine Learning Models. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2023, 538–547. ACM Press.

Stewart, R.; Palmer, T.; and DuPont, B. 2021. A survey of multi-objective optimization methods and their applications for nuclear scientists and engineers. *Progress in Nuclear Energy*, 138: 103830.

Tamssaouet, K.; Dauzère-Pérès, S.; Knopp, S.; Bitar, A.; and Yugma, C. 2022. Multiobjective optimization for complex flexible job-shop scheduling problems. *European Journal of Operational Research*, 296(1): 87–100.

Wietheger, S.; and Doerr, B. 2023. A Mathematical Runtime Analysis of the Non-dominated Sorting Genetic Algorithm III (NSGA-III). In *Proceedings of the International Joint Conference on Artificial Intelligence*, IJCAI 2023, 5657–5665. ijcai.org.

Wietheger, S.; and Doerr, B. 2024. Near-Tight Runtime Guarantees for Many-Objective Evolutionary Algorithms. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVIII, 153–168.

Witt, C. 2014. Fitness levels with tail bounds for the analysis of randomized search heuristics. *Information Processing Letters*, 114: 38–41.

Yadav, D.; Ramu, P.; and Deb, K. 2023. Finding Robust Solutions for Many-Objective Optimization Using NSGA-III. In *2023 IEEE Congress on Evolutionary Computation (CEC)*, 1–8.

Zheng, W.; and Doerr, B. 2022. Better Approximation Guarantees for the NSGA-II by Using the Current Crowding Distance. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2022, 611–619. ACM Press.

Zheng, W.; and Doerr, B. 2024a. Runtime Analysis for the NSGA-II: Proving, Quantifying, and Explaining the Inefficiency for Many Objectives. *IEEE Transactions on Evolutionary Computation*, 28(5): 1442–1454.

Zheng, W.; and Doerr, B. 2024b. Runtime Analysis of the SMS-EMOA for Many-Objective Optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2024, 20874–20882. AAAI Press.

Zheng, W.; Gao, Y.; and Doerr, B. 2024. A Crowding Distance That Provably Solves the Difficulties of the NSGA-II in Many-Objective Optimization. arXiv:2407.17687.

Zheng, W.; Liu, Y.; and Doerr, B. 2022. A First Mathematical Runtime Analysis of the Non-dominated Sorting Genetic Algorithm II (NSGA-II). In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2022, 10408–10416. AAAI Press.