A First Runtime Analysis of NSGA-III on a Many-Objective Multimodal Problem: Provable Exponential Speedup via Stochastic Population Update

Andre Opris

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

Abstract

The NSGA-III is a prominent algorithm in evolutionary many-objective optimization. It is well-suited for optimizing functions with more than three objectives, setting it apart from the classic NSGA-II. However, theoretical insights about NSGA-III of when and why it performs well are still in its early development. This paper addresses this point and conducts a rigorous runtime analysis of NSGA-III on the many-objective ONEJUMPZEROJUMP benchmark (OJZJ for short), providing runtime bounds where the number of objectives is constant. We show that NSGA-III finds the Pareto front of OJZJ in time $O(n^{k+d/2} + \mu n \ln(n))$ where n is the problem size, d is the number of objectives, k is the gap size, a problem specific parameter, if its population size $\mu \in 2^{O(n)}$ is at least $(2n/d + 1)^{d/2}$. Notably, NSGA-III is faster than NSGA-II by a factor of $\mu/n^{d/2}$ for some $\mu \in \omega(n^{d/2})$. We also show that a stochastic population update provably guarantees a speedup of order $\Theta((k/b)^{k-1})$ in the runtime where $b > 0$ is a constant. Apart from a paper of Wietheger and Doerr (PPSN 2024), this is the first rigorous runtime analysis of NSGA-III on OJZJ. Proving these bounds requires a much deeper understanding of the population dynamics of NSGA-III than previous papers achieved.

1 Introduction

Evolutionary multi-objective algorithms (EMOAs) use principles of nature to optimize functions with multiple conflicting objectives by finding a Pareto optimal set [Emmerich and Deutz, 2018]. These have been commonly applied to a wide range of optimization problems in practice [Zapotecas-Martínez *et al.*, 2023; Keller, 2017; Gunantara, 2018] which also include neural networks [Liu *et al.*, 2020], bioinformatics [Handl *et al.*, 2008], engineering [Sharma and Chahar, 2022] or various fields of artificial intelligence [Luukkonen *et al.*, 2023; Zhang *et al.*, 2021; Monteiro and Reynoso-Meza, 2023]. As it is typical for real world problems [Stewart *et al.*, 2008], such problems often involve four or more objectives. Hence, the study

of EMOAs on many-objective problems has quickly gained huge importance in many research fields. However, when the number of objectives increases, the Pareto front grows exponentially in the number of objectives and hence problems typically become more challenging. Additionally, it becomes more difficult to identify dependencies between single objectives. NSGA-II [Deb *et al.*, 2002], the most prominent EMOA ($\sim$53.000 citations), is able to optimize bi-objective problems efficiently (see [Zheng *et al.*, 2022] for a first rigorous analysis and for example [Dang *et al.*, 2023b; Doerr and Qu, 2022] for further rigorous ones or [Deb *et al.*, 2002] for empirical results), but loses performance if the number of objectives grows (see [Zheng and Doerr, 2024a] for rigorous results where there is a huge difference already between two and three objectives or [Khare *et al.*, 2003; Purshouse and Fleming, 2007] for empirical studies). This behaviour comes from the *crowding distance*, the tie breaker in NSGA-II, which is based on sorting search points in each objective. In case of two objectives, a sorting of non-dominated individuals with respect to the first objective induces a sorting with respect to the second one (in reverse order) and hence the crowding distance is a good measure for the closeness of an individual to its neighbors in the objective space. However, for problems with three or more objectives, such a correlation between different objectives does not necessarily exist and hence individuals may have crowding distance zero even if they are not close to others in the objective space. To overcome this problem, the NSGA-III algorithm was designed in [Deb and Jain, 2014] which uses a set of predefined reference points instead of the crowding distance. It can optimize a broad class of different benchmarks with at least four objectives efficiently [Yannibelli *et al.*, 2020; Gu *et al.*, 2022], demonstrating its success in practice ($\sim$6.000 citations). However, theoretical understanding of its success is still in its early development and lags far behind its practical impact. To the best of our knowledge, there are only a few papers which address rigorous runtime analyses of this algorithm [Wietheger and Doerr, 2023; Wietheger and Doerr, 2024; Opris *et al.*, 2024; Opris, 2025] and only [Wietheger and Doerr, 2024] investigated a multimodal problem like the d-OJZJ problem, where the algorithm has to cross a fitness valley to cover the entire Pareto front. But especially on functions with local optima it is important to understand the mechanics of NSGA-III and its variants to

obtain valuable insights when and why it performs well since such many-objective problems occur very often in real world scenarios [Geng *et al.*, 2023]. This may help practitioners to design improved versions of NSGA-III with enhanced performance.

In this paper we significantly extend the results from [Wietheger and Doerr, 2024] and provide a runtime analysis of NSGA-III on d-OJZJ with and without stochastic population update. Our three main contributions are detailed as follows. Firstly, we show an expected runtime bound on NSGA-III on d-OJZJ of $O(n^{k+d/2}/\mu + n\ln(n))$ generations, or $O(n^{k+d/2}+n\ln(n)\mu)$ fitness evaluations, for a constant number d of objectives and population size $\mu \geq (2n/d+1)^{d/2}$, but $\mu \in 2^{O(n)}$ and $2 \leq k \leq n/(2d)$. This is by a factor of $\min\{\mu/n^{d/2}, n^{k-1}/\ln(n)\}$ smaller than the bound of $O(n^k)$ computed in [Wietheger and Doerr, 2024]. Remarkably, for $\mu \in \omega(n^{d/2}) \cap o(n^{k-1}/\ln(n))$, our bound is $o(n^k\mu)$, and hence if $d = 2$, NSGA-III outperforms NSGA-II if also $\mu \in o(n^2/k^2)$ (compare with [Doerr and Qu, 2023a] for the tight runtime bound of $\Theta(\mu n^k)$ of NSGA-II on OJZJ if $n+1 \leq \mu \in o(n^2/k^2)$). We also derive a lower runtime bound of NSGA-III on OJZJ of $\Omega(n^{k+1}/\mu)$ generations, or $\Omega(n^{k+1})$ fitness evaluations in expectation, for $2 \leq k \leq n/4$ and $\mu \in O(n^{k-1})$ when the number of objectives is $d = 2$. This implies tightness if also $n + 1 \leq \mu = O(n^k/\ln(n))$. This is another novel approach, since, as far as we know, the only rigorously proven lower runtime bounds for NSGA-III are provided in [Opris, 2025] on an artificial benchmark quite different to d-OJZJ.

Secondly, to be able to derive these runtime bounds, we have to investigate the population dynamics of NSGA-III more carefully than previous works did. To this end, we use the *cover number* of an objective vector v, which is the number of individuals in the current population with fitness vector v. Outgoing from some general results applicable to any fitness function, we prove with probability $1 - e^{-\Omega(n)}$ that after $O(n)$ generations, when there are only Pareto optimal individuals, the cover number of every Pareto optimal vector is at most $c\mu/n^{d/2}$ for a suitable constant c. This also holds in all future iterations if this property about Pareto optimality is not violated. Hence, all solutions are spread out quite evenly on a large fraction of the Pareto front. For example, if $\mu = \Theta(n^{d/2})$, the cover number of every v is then only at most constant.

Thirdly, we investigate NSGA-III with stochastic population update where the next generation is not only formed by selecting the first ranked solutions resulting from non-dominated sorting, i.e. in a greedy, deterministic manner, but also by selecting some individuals chosen uniformly at random [Bian *et al.*, 2023]. Hence, low-ranked, but promising solutions also have a certain chance to survive. This feature may lead to an exponential speedup in the runtime. This version of NSGA-III can optimize OJZJ in $O(k(bn/k)^k)$ generations, where $b > 1$ is a suitable constant, extending results from [Bian *et al.*, 2023] for NSGA-II and $d = 2$ to NSGA-III and many objectives (see also [Zheng and Doerr, 2024b] for similar results regarding the SMS-EMOA). This is by at least a factor of $\Omega((k/b)^{k-1}n^{d/2}/\mu)$ smaller compared to the upper bound above which is exponentially large if k is linear in n and $\mu = \Theta(n^{d/2})$.

To achieve these, we also have to adapt the arguments from [Opris *et al.*, 2024] about the protection of good solutions also to the case when stochastic population update is considered.

There are several theoretical runtime analyses showcasing the efficiency of NSGA-II on bi-objective problems. The first was conducted by [Zheng *et al.*, 2022] on classical benchmark problems, followed by results on a multimodal problem [Doerr and Qu, 2022], about the usefulness of crossover [Dang *et al.*, 2024a; Doerr and Qu, 2023b], noisy environments [Dang *et al.*, 2023a], approximations of covering the Pareto front [Zheng and Doerr, 2022], lower bounds [Doerr and Qu, 2023a], trap functions [Dang *et al.*, 2024b] and stochastic population update [Bian *et al.*, 2023]. There are also results on combinatorial optimization problems like the minimum spanning tree problem [Cerf *et al.*, 2023] or the subset selection problem [Deng *et al.*, 2024]. However, rigorous runtime results in many-objective optimization on simple benchmark functions appeared only recently for the SMS-EMOA [Zheng and Doerr, 2024b], the SPEA2 [Ren *et al.*, 2024], variants of the NSGA-II [Doerr *et al.*, 2025] and the NSGA-III [Wietheger and Doerr, 2023; Opris *et al.*, 2024]. For the NSGA-III, in [Wietheger and Doerr, 2023] the first runtime analysis on the 3-ONEMINMAX problem with $p \geq 21n$ divisions along each objective was conducted for defining the set of reference points. In [Opris *et al.*, 2024] this result was generalized to more than three objectives and runtime analyses for the classical d-COUNTINGONESCOUNTINGZEROES and d-LEADINGONESTRAILINGZEROES benchmarks [Laumanns *et al.*, 2004] for any constant number d of objectives were provided where it is also necessary to reach the Pareto front. They could also reduce the number of required divisions by more than half. Finally, the first runtime analysis of NSGA-III on d-OJZJ is given in [Wietheger and Doerr, 2024].

2 Preliminaries

For a finite set A we denote by $|A|$ its cardinality and by $\ln$ the logarithm to base e. For $n \in \mathbb{N}$ let $[n] := \{1, \ldots, n\}$. The number of ones in a bit string x is denoted by $|x|_1$ and the number of zeros by $|x|_0$, respectively. For two random variables Y and Z on $\mathbb{N}_0$ we say that Z *stochastically dominates* Y if $P(Z \leq c) \leq P(Y \leq c)$ for every $c \geq 0$. For a d-objective function $f : \{0,1\}^n \to \mathbb{N}_0^d, x \mapsto (f_1(x), \ldots, f_d(x))$, let $f_{\max} := \max\{f_j(x) \mid x \in \{0,1\}^n, j \in [d]\}$ be the maximum possible objective value. When $d = 2$, f is also called *bi-objective*. For two search points $x, y \in \{0,1\}^n$, x *weakly dominates* y, written as $x \succeq y$, if $f_i(x) \geq f_i(y)$ for all $i \in [d]$ and x *(strictly) dominates* y, written as $x \succ y$, if one inequality is strict. We call x and y *incomparable* if neither $x \succeq y$ nor $y \succeq x$. A set $I \subseteq \{0,1\}^n$ is a *set of mutually incomparable solutions* if all search points in I are incomparable. Each solution x not dominated by any other in $\{0,1\}^n$ is called *Pareto-optimal* and we call $f(x)$ *non-dominated fitness value*. The set of all non-dominated fitness values is called *Pareto front*. For a

---

Algorithm 1: NSGA-III ([Deb and Jain, 2014]) with population size μ, stochastic population update ($a = 1$) and without ($a = 0$) on a d-objective function f

---

1 Initialize $P_0 \sim \text{Unif}((\{0,1\}^n)^\mu)$
2 **for** $t := 0$ *to* ∞ **do**
3 $\quad$ Initialize $Q_t := \emptyset$
4 $\quad$ **for** $i = 1$ *to* μ **do**
5 $\quad\quad$ Sample s from P_t uniformly at random
6 $\quad\quad$ Create r by standard bit mutation on s with mutation probability $1/n$
7 $\quad\quad$ Update $Q_t := Q_t \cup \{r\}$
8 $\quad$ Set $R_t := P_t \cup Q_t$
9 $\quad$ **if** $a = 1$ **then**
10 $\quad\quad$ Update R_t by choosing $\lceil 3\mu/2 \rceil$ solutions from R_t uniformly at random without replacement
11 $\quad$ Partition R_t into layers $F_t^1, F_t^2, \ldots, F_t^k$ of non-dominated solutions
12 $\quad$ **if** $a = 0$ **then**
13 $\quad\quad$ Find $i^* \geq 1$ such that $\sum_{i=1}^{i^*-1} |F_t^i| < \mu$ and $\sum_{i=1}^{i^*} |F_t^i| \geq \mu$
14 $\quad$ **else**
15 $\quad\quad$ Find $i^* \geq 1$ such that $\sum_{i=1}^{i^*-1} |F_t^i| < \lceil \mu/2 \rceil$ and $\sum_{i=1}^{i^*} |F_t^i| \geq \lceil \mu/2 \rceil$
16 $\quad$ Compute $Y_t = \bigcup_{i=1}^{i^*-1} F_t^i$
17 $\quad$ Choose $\tilde{F}_t^{i^*} \subset F_t^{i^*}$ such that $|Y_t \cup \tilde{F}_t^{i^*}| = \mu$ if $a = 0$ and $|Y_t \cup \tilde{F}_t^{i^*}| = \lceil \mu/2 \rceil$ if $a = 1$ with Algorithm 2
18 $\quad$ Create the next population $P_{t+1} := Y_t \cup \tilde{F}_t^{i^*}$ if $a = 0$ and $P_{t+1} := Y_t \cup \tilde{F}_t^{i^*} \cup ((P_t \cup Q_t) \setminus R_t)$ if $a = 1$

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

---

1 Compute the normalisation f^n of f
2 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
3 For each $r \in \mathcal{R}_p$, set $\rho_r := |\{x \in Y_t \mid \text{rp}(x) = r\}|$
4 Initialize $\tilde{F}_t^{i^*} = \emptyset$ and $R' := \mathcal{R}_p$
5 **while** *true* **do**
6 $\quad$ Determine $r_{\min} \in R'$ such that $\rho_{r_{\min}}$ is minimal (where ties are broken randomly)
7 $\quad$ 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)
8 $\quad$ **if** $x_{r_{\min}}$ *exists* **then**
9 $\quad\quad$ $\tilde{F}_t^{i^*} = \tilde{F}_t^{i^*} \cup \{x_{r_{\min}}\}$
10 $\quad\quad$ $\rho_{r_{\min}} = \rho_{r_{\min}} + 1$
11 $\quad\quad$ **if** $a = 0$ and $|Y_t| + |\tilde{F}_t^{i^*}| = \mu$ **then**
12 $\quad\quad\quad$ **return** $\tilde{F}_t^{i^*}$
13 $\quad\quad$ **if** $a = 1$ and $|Y_t| + |\tilde{F}_t^{i^*}| = \lceil \mu/2 \rceil$ **then**
14 $\quad\quad\quad$ **return** $\tilde{F}_t^{i^*}$
15 $\quad$ **else** $R' = R' \setminus \{r_{\min}\};$

---

population P_t the *cover number* $c_t(v)$ of $v \in \mathbb{N}_0^d$ is the number of individuals $x \in P_t$ with $f(x) = v$ and we say that v is *covered* if its cover number is at least 1.

The NSGA-III algorithm, originated in [Deb and Jain, 2014], with or without *stochastic population update* 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. When stochastic population update is turned on (i.e. $a = 1$) then R_t is updated by choosing $\lceil 3\mu/2 \rceil$ individuals from R_t uniformly at random without replacement. Then R_t is partitioned into layers $F_{t+1}^1, F_{t+1}^2, \ldots$ 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 $F_{t+1}^1, \ldots, F_{t+1}^{i-1}$. Then if $a = 0$ the critical rank i^* with $\sum_{i=1}^{i^*-1} |F_i^t| < \mu$ and $\sum_{i=1}^{i^*} |F_i^t| \geq \mu$

is determined (i.e. there are fewer than μ search points in R_t with a lower rank than i^*, but at least μ search points with rank at most i^*). If $a = 1$, then every individual not selected for R_t survives. The number of such individuals is $\mu - \lceil 3\mu/2 \rceil = \lfloor \mu/2 \rfloor$. Hence, in this case, the critical index i^* refers only on $\mu - \lfloor \mu/2 \rfloor = \lceil \mu/2 \rceil$ individuals. All individuals with a lower rank than i^* are included in P_{t+1}, while the remaining individuals are selected from $F_{i^*}^t$ 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 a d-objective function $f : \{0,1\}^n \to \mathbb{N}_0^d$, the normalized fitness vector $f^n(x) := (f_1^n(x), \ldots, f_d^n(x))$ 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 $j \in [d]$ where $y^{\text{nad}} := (y_1^{\text{nad}}, \ldots, y_d^{\text{nad}})$ and $y^{\min} := (y_1^{\min}, \ldots, y_d^{\min})$ 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 $j \in [d]$ where ε_{nad} is a positive threshold set by the user (see [Blank et al., 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 $P_0, Q_0, \ldots, P_t, Q_t$). After computing the nor-
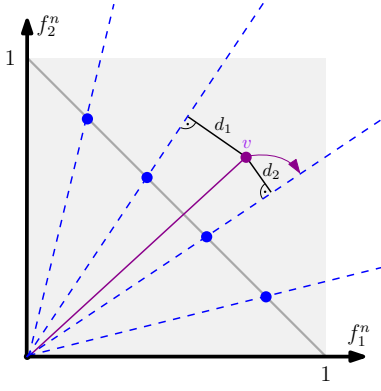
Figure 1: Illustrating how search points with fitness vector v are associated with reference points (dots on line through $(1,0)$ and $(0,1)$ connected by dashed lines through the origin) for $m = 2$ objectives in the normalized objective space. The vector v is associated to the nearest reference point to its right.

malisation, each individual x is associated with the reference point $\mathrm{rp}(x)$ such that the distance between $f^n(x)$ and the line through the origin and $\mathrm{rp}(x)$ is minimal (see also Figure 1). Ties are broken deterministically, which means that two or more individuals which have the same smallest distance to two or more reference points are associated to the same reference point. We use the same set of reference points $\mathcal{R}_p$ as proposed in [Deb and Jain, 2014]. The points are defined as

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

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 $(1, 0, \ldots, 0)^\top, (0, 1, \ldots, 0)^\top, \ldots, (0, 0, \ldots, 1)^\top$.
Then, one iterates through all the reference points where the reference point with the fewest associated individuals that are already selected for the next generation P_{t+1} is chosen. A reference point is omitted if it only has associated individuals that are already selected for P_{t+1} and ties are broken uniformly at random. Next, from the individuals associated to that reference point who have not yet been selected, the one closest to the chosen reference point is selected for the next generation, where ties are again broken uniformly at random. Once the required number of individuals is reached (i.e. if $|Y_t| + |\tilde{F}_t^{i^*}| = \mu$ when $a = 0$ and $|Y_t| + |\tilde{F}_t^{i^*}| = \lceil \mu/2 \rceil$ otherwise) the selection ends.

3 Structural Results

The following result from [Opris $et\ al.$, 2024] also adapted to the case when stochastic population update is turned on (i.e. $a = 1$) shows that if a population covers a fitness vector v with a first-ranked individual x then it is covered for all future generations as long as $f(x)$ is non-dominated. In other words, the best solutions are protected.

Lemma 1. *Consider NSGA-III on a d-objective function f with $\varepsilon_{nad} \geq f_{\max}$ and a set $\mathcal{R}_p$ of reference points for $p \in \mathbb{N}$*

with $p \geq 2d^{3/2} f_{\max}$. Let P_t be its current population. Assume $\mu \geq |I|$ if $a = 0$ or $\mu \geq 2|I|$ otherwise for the population size μ where I is a maximum set of mutually incomparable solutions. Let $\tilde{F}_t^1, \tilde{F}_t^2, \ldots$ be the layers of non-dominated fitness vectors of $P_t \cup Q_t$. Then for every $x \in \tilde{F}_t^1$ there is an $x' \in P_{t+1}$ with $f(x') = f(x)$.

Proof. The case $a = 0$ is Lemma 3.4 in [Opris $et\ al.$, 2024]. So suppose that $a = 1$. Let $x \in \tilde{F}_t^1$. If $x \notin R_t$ then the claim holds (by Line 18 in Algorithm 1 we have $(P_t \cup Q_t) \setminus R_t \subset P_{t+1}$). Suppose that $x \in R_t$. Then $x \in F_t^1$ where F_t^1 is the first layer of non-dominated solutions from Line 11 in Algorithm 1 (since if x is non-dominated with respect to $P_t \cup Q_t$ then it is also non-dominated with respect to the updated R_t from Line 10 in Algorithm 1). If $|F_t^1| \leq \lceil \mu/2 \rceil$ all individuals in F_t^1 (including x) survive. Hence, suppose that $|F_t^1| > \lceil \mu/2 \rceil$ and let $s := |\{f(x) \mid x \in F_t^1\}|$ be the number of different fitness vectors of search points from F_t^1. Then by Lemma 3.3 in [Opris $et\ al.$, 2024] individuals with distinct fitness vectors in F_t^1 are associated with different reference points. Thus there are s reference points, each associated with at least one individual. Since $\lceil \mu/2 \rceil \geq |I| \geq s$, at least one individual $x' \in R_t$ associated with the same reference point as x survives and we have $f(x) = f(x')$. $\square$

The population dynamics of NSGA-III significantly differs to that of NSGA-II. In case of NSGA-II, which uses the crowding distance as second tie breaker, all rank one individuals with crowding distance zero are treated equally in the survival selection no matter how they are distributed. In contrast, NSGA-III (without stochastic population update) keeps uniform distributions of Pareto optimal individuals across the search space. This behaviour can be formally described using the *cover number* $c_t(v)$ of a fitness vector v, defined as the number of individuals in P_t with fitness v, captured in the following lemma.

Lemma 2. *Assume the same conditions as in Lemma 1 where stochastic population update is disabled. The following properties hold.*

(1) Let $v \in F^*$ and $0 \leq a \leq \mu/|I|$. If $c_t(v) \geq a$ then $c_{t+1}(v) \geq a$.

(2) 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 $w \in F^*$.

(3) Suppose that every $x \in P_t$ is Pareto optimal. Then $m_t := \max\{c_t(v) \mid v \in F^*\}$ does not increase.

Proof. (1): The NSGA-III iterates through all reference points, always preferring a reference point r with the fewest associated individuals chosen for P_{t+1} so far (see Line 6 in Algorithm 2), and selecting an individual for P_{t+1} associated to r. Since, by Lemma 3.3 in [Opris $et\ al.$, 2024], two Pareto optimal search points with distinct fitness are associated to two different reference points, NSGA-III iterates at least a times through all reference points with at least a many associated individuals to find P_{t+1}. Hence, the cover number $c_{t+1}(v)$ of v with respect to P_{t+1} is still at least a.
(2): There is a reference point r to that all x with $f(x) = v$ are associated. Hence, NSGA-III iterates at most $c_t(v) - 1$

times through r, since $c_{t+1}(v) < c_t(v)$, and consequently, through all reference points at most $c_t(v)$ times (see Line 6 in Algorithm 2). Thus, for every $w \in V$, at most $c_t(v)$ individuals x with $f(x) = w$ survive.

(3): By iterating through all reference points at most m_t times, the algorithm finds μ individuals in F_t^1 (since all Pareto optimal individuals are in F_t^1) and the survival selection ends. Then $c_{t+1}(w) \leq m_t$ for every $w \in F^*$. $\qquad\square$

4 The Many-Objective Jump Function

In this section we define the d-ONEJUMPZEROJUMP$_k$ (d-OJZJ$_k$) function, defined in [Zheng and Doerr, 2024b], as a d-objective version of the bi-objective JUMP$_k$ benchmark. The latter was introduced in [Doerr and Qu, 2022] to understand how MOEAs can handle functions with local optima, i.e. changes of size $k \geq 2$ are necessary to cover the whole Pareto front. Fix $d \in \mathbb{N}$ divisible by two and let n be divisible by $d/2$. For a bit string x let $x := (x^1, \ldots, x^{d/2})$ where all x^j are of equal length $2n/d$. For odd $j \in [d]$ define $y^j := x^{(j+1)/2}$ and $y^j := x^{j/2}$ for even $j \in [d]$. For $2 \leq k \leq 2n/d$ the d-OJZJ$_k(x) = (f_1(x), \ldots, f_d(x))$ is defined as

$$f_j(x) = \begin{cases} k + |y^j|_1, & \text{if } |y^j|_1 \leq \frac{2n}{d} - k \text{ or } y^j = 1^{2n/d}, \\ \frac{2n}{d} - |y^j|_1, & \text{else,} \end{cases}$$

if $j \in [1, \ldots, d]$ is odd, and

$$f_j(x) = \begin{cases} k + |y^j|_0, & \text{if } |y^j|_0 \leq \frac{2n}{d} - k \text{ or } y^j = 0^{2n/d}, \\ \frac{2n}{d} - |y^j|_0, & \text{else,} \end{cases}$$

if $j \in [1, \ldots, d]$ is even. Note that for odd $j \in [d]$ the $(j + 1)$-th objective is structurally identical to the j-th one with the roles of ones and zeros reversed. For every objective there are $2n/d + 1$ different values and the maximum possible value is $k + 2n/d$. In [Zheng and Doerr, 2024b] it is shown that the Pareto front F^* of d-OJZJ$_k$ is $\{(a_1, 2k + 2n/d - a_1, \ldots, a_{d/2}, 2k + 2n/d - a_{d/2}) \mid a_1, \ldots, a_{d/2} \in \{k, 2k, 2k+1, \ldots, 2n/d-1, 2n/d, 2n/d+k\}\}$ and has cardinality $(2n/d-2k+3)^{d/2}$ for $k \leq n/d$. Further, a maximum set of mutually incomparable solutions I (its cardinality can be larger than $|F^*|$, compare with [Zheng and Doerr, 2024b]) fulfills $|F^*| \leq |I| \leq (2n/d + 1)^{d/2}$ since I does not contain two search points with the same fitness vector and hence $|I| \leq |f(\{0,1\}^n)| = (2n/d + 1)^{d/2}$. For $d = 2$ we have that every non-Pareto optimal individual is dominated by a Pareto optimal one which is different in the many-objective case.

Lemma 3. *Suppose that $d = 2$ and let y be not Pareto optimal. Then every Pareto optimal x strictly dominates y.* [1]

To simplify the analysis, we assume that d is constant.

5 Population Dynamics of NSGA-III on OJZJ

In this section we investigate the population dynamics of the steady state NSGA-III on d-OJZJ and demonstrate that, in

---

[1]A long version of this paper can be found at https://arxiv.org/abs/2505.01256 where the reader can find proofs for all lemmas stated in the paper.

an expected linear number of generations, the population is evenly distributed across a subset $V \subset F^*$ already covered by P_t. Furthermore, if the population P_t is Pareto optimal in generation t and all future ones, the distribution can be even more characterized by bounding the cover number of every $v \in F^*$ by $\lceil \mu/|V| \rceil$ from above. Notably, for $|V| = |F^*|$, this is the best possible distribution of P_t across the whole Pareto front one can achieve. We also demonstrate that once such a distribution is reached, it is maintained for all future generations.

Lemma 4. *Assume the same conditions as in Lemma 1 where stochastic population update is disabled. Suppose $\mu \in 2^{O(n)}$. Let $V \subset F^*$ be covered by P_t. The following properties are satisfied after $O(n)$ generations and all future ones with probability at least $1 - e^{-\Omega(n)}$ and in expectation.*

(1) For all $v \in V$ we have $c_t(v) \geq \beta := \lfloor \mu/(2n/d + 1)^{d/2} \rfloor$.

(2) If all populations $P_0, P_1, \ldots$ consist only of Pareto optimal individuals, we have that $c_t(v) \leq \lceil \mu/|V| \rceil$.

Proof. (1): Fix a Pareto optimal x with $v := f(x) \in V$. We show that $c_t(v) \geq \beta$ with probability at least $1 - e^{-\Omega(n)}$ after $O(n)$ generations. By Lemma 2(1), $c_t(v) \leq \beta$ cannot decrease since a maximum set of mutually incomparable solutions I fulfills $|I| \leq (2n/d + 1)^{d/2}$ and hence, $c_t(v) \leq \beta \leq \mu/|I|$. We define two phases where the second phase only applies if $\beta > n$.
Phase 1. We have $c_t(v) \geq \ell := \min\{\beta, n\}$.
Phase 2. We have $c_t(v) \geq \beta$.
In the next lemma we determine the expected duration of each phase under the assumptions of Lemma 4.

Lemma 5. *Both phases are finished in $O(n)$ generations with probability at least $1 - e^{-\Omega(n)}$ and in expectation.*

Now we obtain by a union bound on both phases and every $v \in V$ that in $O(n)$ generations the cover number of every $v \in V$ is at least β with probability at least $1 - e^{-\Omega(n)}$ (since $|V| = O(n^{d/2}) = e^{o(n)}$). The bound on the expected number of generations follows by applying the same arguments for an additional period of $O(n)$ generations and by the fact that $(1 + o(1))O(n) = O(n)$ such periods are sufficient.
(2): With the same argument as in (1) we obtain with probability $1 - e^{-\Omega(n)}$ that, after $O(n)$ generations, the cover number of all $v \in V$ is at least $\lceil \mu/|V| \rceil$ or one of these cover numbers has decreased at least one time when it was at most $\lceil \mu/|V| \rceil$. Suppose that the former happens. If $\lceil \mu/|V| \rceil > \mu/|V|$ we see that $\lceil \mu/|V| \rceil \cdot |V| > |V| \cdot \mu/|V| = \mu$, a contradiction. Hence, $\lceil \mu/|V| \rceil = \mu/|V|$ and all μ individuals are completely evenly distributed on V. If the latter happens, we see with Lemma 2(2) that the cover number of all vectors $v \in F^*$ is at most $\lceil \mu/|V| \rceil$. By Lemma 2(3) the maximum cover number cannot increase, proving Lemma 4. $\qquad\square$

6 Upper Runtime Bounds

In this section we prove upper runtime bounds for NSGA-III with stochastic population update ($a = 1$), as well as without ($a = 0$), on $f := d$-OJZJ. For the variant where $a = 1$ we

follow the proof in [Bian *et al.*, 2023]. For technical reasons, we define for a Pareto optimal fitness vector $v = f(x)$ the string $L(v) \in \{0, 1, \perp\}^d$ as follows. For $j \in [d]$ let

- $L(v)_j = 1$ if $v_j = 2n/d + k$ and j is odd (attained if $x^{(j+1)/2} = 1^n$ since then $f_j(x) = 2n/d + k$),
- $L(v)_j = 0$ if $v_j = 2n/d + k$ and j is even (attained if $x^{j/2} = 0^n$ since then $f_j(x) = 2n/d + k$), and
- $L(v)_j = \perp$ if $k \leq v_j \leq 2n/d$ (attained if $k \leq |x^{(j+1)/2}|_1 \leq 2n/d - k$ or $x = 0^{2n/d}$ if j is odd and $k \leq |x^{j/2}|_1 \leq 2n/d - k$ or $x = 1^{2n/d}$ if j is even).

Note that for every Pareto optimal x there is $w \in \{0, 1, \perp\}^d$ with $w = L(f(x))$.

Theorem 6. *Assume the same conditions as in Lemma 1 for $f := d\text{-OJZJ}_k$ where d is constant, $2 \leq k \leq n/(2d)$ and $\mu \in 2^{O(n)}$. Then the expected number of generations until the whole Pareto front is covered is at most $O(n^{k+d/2}/\mu + n \ln(n))$ for NSGA-III if $a = 0$, and $O(k(\frac{12en}{k})^k)$ for NSGA-III if $a = 1$.*

Proof. We fix $v \in F^*$ with string $L(v) \in \{0, 1, \perp\}^d$ and show that all Pareto optimal $w \in F^*$ with $L(w) = L(v)$ are covered after $O(n^{k+d/2}/\mu + n \ln(n))$ generations for NSGA-III if $a = 0$ and after $O(n^{k+d/2}/\mu + n \ln(n))$ generations for NSGA-III if $a = 1$. Since there are at most 3^d different strings $L(v) \in \{0, 1, \perp\}^d$, the runtime follows since $d = \Theta(1)$. Let $O_v := \{i \in [d] \mid (L(v))_i = 1\}$ and $Z_v := \{i \in [d] \mid (L(v))_i = 0\}$. By a classical Chernoff bound, the probability is at most $e^{-\Omega(\mu n)}$ that for all initialized x there is a block $j \in [d/2]$ with $|x^j|_1 \notin \{k, \dots, 2n/d - k\}$. Suppose that this happens. Then the probability is at least n^{-n} to create any individual with mutation. Hence, the expected number of generations to create a Pareto optimal x with $O_{f(x)} = Z_{f(x)} = \emptyset$ (that is, $L(f(x)) = \{\perp\}^d$) is $1 + n^{-n}e^{-\Omega(\mu n)} = 1 + o(1)$. Now, suppose that there is a Pareto optimal x (that is, $L(f(x)) \in \{0, 1, \perp\}^d$) with $O_{f(x)} \subsetneq O_v$ or $Z_{f(x)} \subsetneq Z_v$. Let $i \in O_v \setminus O_{f(x)}$. Note that i is odd. Now we determine the time until all Pareto optimal fitness vectors w with $L(w) = L(f(x))$ are covered, captured by the following lemma where the same assumptions as in Theorem 6 hold.

Lemma 7. *Let $z \in P_t$ be Pareto optimal. Then in $8dn\ln(n)$ generations (i.e. $8\mu dn\ln(n)$ fitness evaluations) all Pareto optimal v with $L(v) = L(f(z))$ are covered with probability at least $1 - n^{-d}$. The expected number of generations is $O(n\ln(n))$.*

To generate a Pareto optimal y with $Z_{f(y)} = Z_{f(x)}$ and $O_{f(y)} = O_{f(x)} \cup \{i\}$ we consider both versions of NSGA-III separately.

For NSGA-III without stochastic population update an additional phase of expected $O(n)$ generations ensures that every fitness vector $w \in F^*$ with $L(w) = L(f(x))$ has cover number at least $c := \lfloor \mu/(2n/d + 1)^{d/2} \rfloor$ (by Lemma 4(1)). Then, to generate y, one can choose z with $L(f(z)) = L(f(x))$ and $|z^{(i+1)/2}|_1 = 2n/d - k$ as a parent (prob. c/μ), and flip k specific bits while keeping the remaining bits unchanged (prob.

$(1 - 1/n)^{n-k}/n^k \geq 1/(en^k)$). Hence, in one generation this happens with probability at least

$$1 - \left(1 - \frac{c}{e\mu n^k}\right)^\mu \geq \frac{c/(en^k)}{1 + c/(en^k)} = \Omega(c/n^k).$$

where the first inequality is due to Lemma 10 in [Badkobeh *et al.*, 2015]. Hence, in total, we obtain $O(n^k/c + n\ln(n)) = O(n^{k+d/2}/\mu + n\ln(n))$ generations in expectation to generate y.

For NSGA-III with stochastic population update we consider a sequence of k successive generations and for $\ell \in [k]$ we call the ℓ-th generation *successful* if an individual z is created with $2n/d - k + \ell$ ones in block $(i+1)/2$ while $z^j = x^j$ for $j \in [d/2] \setminus \{(i+1)/2\}$ and z is not removed. Suppose that the $(\ell - 1)$-th generation is successful. Then, generation ℓ is successful if in one trial one chooses a parent p with $|p^{(i+1)/2}|_1 = 2n/d - k + \ell - 1$ and $z^j = x^j$ for $j \in [d/2] \setminus \{(i+1)/2\}$, flips a zero bit in the $((i+1)/2)$-th block while keeping the remaining bits unchanged and finally keeps the new created individual. With probability at least

$$1 - \left(1 - \frac{k - \ell + 1}{e\mu n}\right)^\mu \geq \frac{\frac{k-\ell+1}{en}}{1 + \frac{k-\ell+1}{en}} \geq \frac{k - \ell + 1}{4n}$$

a desired individual z is created in generation ℓ. Note that z is not Pareto-optimal for $\ell < k$. However, $\mu - \lceil 3\mu/2 \rceil$ individuals chosen uniformly at random survive (see Line 10 and Line 18 in Algorithm 1 for $a = 1$) and hence even a non Pareto optimal individual remains with probability at least $(\mu - \lceil 3\mu/2 \rceil)/\mu = \lfloor \mu/2 \rfloor/\mu \geq (\mu/2 - 1)/\mu = 1/2 - 1/\mu$. For n sufficiently large, we have $1/2 - 1/\mu \geq 1/3$ and hence with probability at least $(k - \ell + 1)/(12n)$, generation ℓ is successful. If all k generations are successful, the desired y is created, which happens with probability at least

$$\prod_{\ell=1}^{k} \frac{k - \ell + 1}{12n} = \frac{k!}{(12n)^k} \geq \frac{k^k}{(12en)^k}$$

(since $k! \geq (k/e)^k$ by Stirling's approximation). Hence, the expected number of generations to create the desired y is at most $k(12en)^k/k^k$.

In either case, by symmetry, the above bounds also apply to create a Pareto optimal y such that $Z_{f(y)} = Z_{f(x)} \cup \{i\}$ for $i \in Z_v \setminus Z_{f(x)}$ and $O_{f(y)} = O_{f(x)}$ else. Since O_v and Z_v are both finite, we obtain that a Pareto optimal individual z with $O_{f(z)} = O_v$ and $Z_{f(z)} = Z_v$ is created in expected $O(n^{k+d/2}/\mu + n\ln(n))$ generations when stochastic population update is turned off and $O(k(\frac{12en}{k})^k)$ it it is turned on. With Lemma 7 we obtain that all Pareto optimal v with $L(v) = L(f(z))$ are covered in expected $O(n\ln(n))$ generations, proving the theorem. □

We see that for $\mu \in \Omega(n^{d/2}) \cap O(n^{k+d/2-1}/\ln(n))$, the derived runtime if $a = 0$ is $O(n^{k+d/2})$ in terms of fitness evaluations and, hence, independent of the population size μ. An important consequence is that, if the number of objectives is 2, NSGA-III outperforms NSGA-II if also $\mu \in \omega(n^{d/2}) \cap o(n^2/k^2)$ (compare with [Doerr and Qu, 2023a] for the tight runtime bound of $\Theta(\mu n^k)$ of NSGA-II on 2-OJZJ in terms of fitness evaluations if $n + 1 \leq \mu \in o(n^2/k^2)$).

7 Lower Runtime Bounds

In this section we prove sharp lower runtime bounds of NS-GA-III without stochastic population update on OJZJ for the bi-objective case which are tight for many parameter settings. To this purpose, we show that with probability $1 - e^{-\Omega(n)}$ after $O(n)$ generations, the cover number of every $v \in F^*$ can be bounded from above by a sufficiently small value, which also remains for all future iterations. In other words, the population remains well-distributed across F^*.

Lemma 8. *Assume the same conditions as in Lemma 1 for $f := d$-OJZJ, $a = 0$, d is constant, $2 \leq k \leq n/(2d)$ and $\mu \in 2^{O(n)}$. Suppose that all $P_0, P_1, \ldots$ ever seen by NSGA-III consist only of Pareto optimal individuals. Then, with probability at least $1 - e^{-\Omega(n)}$ the cover number of each Pareto optimal fitness vector is at most $\lceil 2^{d/2+1} \mu/|I| \rceil$ after $O(n)$ generations.*

However, in the case $d > 2$ it may happen that non-Pareto optimal search points with rank one are created which survive (for example if, outgoing from a population on the Pareto front without $0^{2n/d}$ or $1^{2n/d}$ in each block $j \in [d/2]$, $\mu - 1$ individuals are cloned and one non-Pareto optimal individual y is generated with $y^i = 1^{2n/d}$ in one block $i \in [d/2]$ while $1 \leq |y^j|_1 < k$ in another block j). Such non-Pareto optimal solutions can remain for a long time in the population and hence, Lemma 2(3) cannot be applied from scratch, complicating the analysis. Hence, we present only the case $d = 2$.

Theorem 9. *Assume the same conditions as in Lemma 1 for $f := d$-OJZJ and $a = 0$ where $d = 2$, $2 \leq k \leq n/4$ and $\mu \in O(n^{k-1}) \cap poly(n)$. Then the expected number of generations required for covering the entire Pareto front is at least $\Omega(n^{k+1}/\mu)$.*

Proof. By a classical Chernoff bound, with probability $1 - e^{-\Omega(n)}$, every individual x fulfills $|x|_1 \in \{k, \ldots, n-k\}$ after initialization. Suppose that this happens. Then there are only Pareto optimal individuals in P_0 and all future populations $P_1, P_2, \ldots$ since a non-Pareto optimal individual is dominated by every Pareto optimal one (see Lemma 3). Since there are μ Pareto optimal individuals in R_t, such a y never survives (due to the non-dominated sorting procedure). Note that $|I| = |F^*|$ (since $|F^*| = n - 2k + 3 \geq n/2 + 3$, f attains $n + 1$ different values in each objective and every non-Pareto optimal individual is dominated by a Pareto optimal one). By Lemma 8 applied on $d = 2$, with probability at least $1 - e^{-\Omega(n)}$ the cover number of every $v \in F^*$ is at most $\lceil 4\mu/|I| \rceil$ after at most $O(n)$ generations. Suppose that this happens. Let n be sufficiently large such that $\mu \leq cn^{k-1}$ for a constant $c > 0$ and the event from Lemma 8 applies on at most δn generations for a further constant $\delta > 0$. Further, suppose that $1/n + 5n/|F^*| \leq \alpha$ for a suitable constant $\alpha > 0$ (due to $|F^*| \in \Theta(n)$). We consider a phase of at most δn generations (i.e. $\mu \delta n$ trials) and show that no individual $y \in \{0^n, 1^n\}$ is generated within $\mu \delta n$ trials with at least constant probability. Since in one trial one has to flip at least k specific one bits or k specific zero bits to create y, the probability that this happens in $\mu \delta n$ trials is at most $1 - (1 - 2/n^k)^{\mu \delta n} \leq 1 - (1 - 2/n^k)^{c\delta n^k} \leq 1 - (1/16)^{c\delta} =: b$

where we used $(1 - 2/n)^n \geq ((1 - 1/n)^n)^2 \geq 1/16$. Note that $b < 1$ is a constant. Hence, we obtain with probability at least $1 - b - e^{-\Omega(n)} = \Omega(1)$ that there is a generation t where every $v \in F^*$ has cover number at most $\lceil 4\mu/|F^*| \rceil$, and every individual $x \in P_t$ fulfills $x \notin \{1^n, 0^n\}$. Suppose that this happens. We estimate the expected time to create $x = 1^n$ (which is the only Pareto optimal search point with fitness $f_1(x) = 2n/d + k$) from above. Note that for a given $k \leq \ell \leq 2k - 1$ there are at most $\lceil 4\mu/|F^*| \rceil \leq 5\mu/|F^*|$ different individuals $y \in P_t$ with *Hamming distance* $H(1^n, y) = \ell$ to 1^n. Note also it requires to flip at least $2k$ specific bits to create 1^n from a y with $H(1^n, y) \geq 2k$. Since $k \geq 2$, all these considerations yield a probability of at most

$$\left(\frac{1}{n}\right)^{2k} + \sum_{j=0}^{k-1} \frac{5}{|F^*|} \cdot \left(\frac{1}{n}\right)^{j+k} \cdot \left(1 - \frac{1}{n}\right)^{n-j-k}$$

$$\leq \left(\frac{1}{n}\right)^{2k} + \frac{5}{|F^*|n^k} \leq \left(\frac{1}{n}\right)^{k+1} \left(\frac{1}{n} + \frac{5n}{|F^*|}\right) \leq \frac{\alpha}{n^{k+1}}$$

to create 1^n in one trial. Hence, in one generation we obtain by a union bound on μ trials that 1^n is created with probability at most $\alpha\mu(1/n)^{k+1}$. Hence, the expected number of generations to create 1^n is at least $(1 - b - e^{-\Omega(n)})n^{k+1}/(\alpha\mu) = \Omega(n^{k+1}/\mu)$ which proves the theorem. $\square$

Combining Theorems 6 and 9, we obtain for $|I| = |F^*| \leq \mu \in O(n^{k-1}) \cap \text{poly}(n)$, $d = 2$ and $2 \leq k \leq n/4$ the tight runtime of $\Theta(n^{k+1}/\mu)$ for NSGA-III on 2-OJZJ.

8 Conclusions

In this paper we provided new insights in the population dynamics of NSGA-III, demonstrating that it quickly spreads its solutions evenly across the whole Pareto front. Our methods were developed with a level of generality that makes them applicable to a broad range of problems. Subsequently, we derived upper runtime bounds for NSGA-III on the ONEJUMPZEROJUMP benchmark, and, for certain regimes of gap size k and population size μ, even tight runtime bounds when the number of objectives is two. Additionally, we showed that stochastic population update, where solutions for survival are not always selected deterministically, can lead to an exponential speedup in runtime. We hope that the techniques developed mark a significant step forward in understanding the behavior of NSGA-III, while also shedding light on its strengths and limitations. To deepen this understanding, future work could focus on the usefulness of stochastic population update on more complex problems with many local optima like the many-objective REALROYALROAD function, proposed by [Opris, 2025], where the algorithm has to cross a large fitness valley. Another research direction could be deriving rigorous lower bounds for NSGA-III on d-OJZJ for $d > 2$, or on further classical problems such as OMM, COCZ, and LOTZ. We also hope that the gained deeper theoretical understanding of the dynamics of NSGA-III on OJZJ will offer useful insights for practitioners, such as facilitating the creation of refined versions of the algorithm, particularly to enhance performance in real world scenarios where many local optima occur.

Acknowledgements

References

[Badkobeh *et al.*, 2015] Golnaz Badkobeh, Per Kristian Lehre, and Dirk Sudholt. Black-box complexity of parallel search with distributed populations. In *Proceedings of the Foundations of Genetic Algorithms*, FOGA 2015, pages 3–15. ACM Press, 2015.

[Bian *et al.*, 2023] Chao Bian, Yawen Zhou, Miqing Li, and Chao Qian. 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, pages 5513–5521. IJCAI Organization, 2023.

[Blank *et al.*, 2019] Julian Blank, Kalyanmoy Deb, and Proteek Chandan Roy. Investigating the normalization procedure of NSGA-III. In *Evolutionary Multi-Criterion Optimization*, pages 229–240. Springer, 2019.

[Cerf *et al.*, 2023] Sacha Cerf, Benjamin Doerr, Benjamin Hebras, Yakob Kahane, and Simon Wietheger. The first proven performance guarantees for the non-dominated sorting genetic algorithm II (NSGA-II) on a combinatorial optimization problem. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, IJCAI 2023, pages 5522–5530. IJCAI Organization, 2023.

[Dang *et al.*, 2023a] Duc-Cuong Dang, Andre Opris, Bahare Salehi, and Dirk Sudholt. Analysing the robustness of NSGA-II under noise. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2023, pages 642–651. ACM Press, 2023.

[Dang *et al.*, 2023b] Duc-Cuong Dang, Andre Opris, Bahare Salehi, and Dirk Sudholt. A proof that using crossover can guarantee exponential speed-ups in evolutionary multi-objective optimisation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2023, pages 12390–12398. AAAI Press, 2023.

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

[Dang *et al.*, 2024b] Duc-Cuong Dang, Andre Opris, and Dirk Sudholt. Illustrating the efficiency of popular evolutionary multi-objective algorithms using runtime analysis. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2024, page 484–492. ACM Press, 2024.

[Deb and Jain, 2014] Kalyanmoy Deb and Himanshu Jain. 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, 2014.

[Deb *et al.*, 2002] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[Deng *et al.*, 2024] Renzhong Deng, Weijie Zheng, Mingfeng Li, Jie Liu, and Benjamin Doerr. 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, page 264–279. Springer, 2024.

[Doerr and Qu, 2022] Benjamin Doerr and Zhongdi Qu. 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, pages 399–412. Springer, 2022.

[Doerr and Qu, 2023a] Benjamin Doerr and Zhongdi Qu. 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, pages 12408–12416. AAAI Press, 2023.

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

[Doerr *et al.*, 2025] Benjamin Doerr, Tudor Ivan, and Martin S. Krejca. Speeding up the NSGA-II with a simple tie-breaking rule. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2025, pages 26964–26972. AAAI Press, 2025.

[Doerr, 2019] Benjamin Doerr. Analyzing randomized search heuristics via stochastic domination. *Theoretical Computer Science*, 773:115–137, 2019.

[Emmerich and Deutz, 2018] Michael T. M. Emmerich and André H. Deutz. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Natural Computing*, 17:585–609, 2018.

[Geng *et al.*, 2023] Huantong Geng, Zhengli Zhou, Junye Shen, and Feifei Song. A dual-population-based NSGA-III for constrained many-objective optimization. *Entropy*, 25(1), 2023.

[Gu *et al.*, 2022] Qinghua Gu, Qingsong Xu, and Xuexian Li. An improved NSGA-III algorithm based on distance dominance relation for many-objective optimization. *Expert Systems with Applications*, 207:117738, 2022.

[Gunantara, 2018] Nyoman Gunantara. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1):1502242, 2018.

[Handl *et al.*, 2008] Julia Handl, Simon C. Lovell, and Joshua Knowles. Investigations into the effect of multiobjectivization in protein structure prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5199:702–711, 2008.

[Keller, 2017] Andre A. Keller. *Multi-Objective Optimization in Theory and Practice I: Classical Methods*. Bentham Science Publishers, 2017.

[Khare *et al.*, 2003] V. Khare, X. Yao, and K. Deb. Performance scaling of multi-objective evolutionary algorithms. In *Evolutionary Multi-Criterion Optimization*, pages 376–390. Springer, 2003.

[Laumanns *et al.*, 2004] Marco Laumanns, Lothar Thiele, and Eckart Zitzler. Running time analysis of multiobjective evolutionary algorithms on pseudo-boolean functions. *IEEE Transactions on Evolutionary Computation*, 8(2):170–182, 2004.

[Liu *et al.*, 2020] Xuning Liu, Yong Zhou, Jiaqi Zhao, Rui Yao, Bing Liu, Ding Ma, and Yi Zheng. Multiobjective resnet pruning by means of EMOAs for remote sensing scene classification. *Neurocomputing*, 381:298–305, 2020.

[Luukkonen *et al.*, 2023] Sohvi Luukkonen, Helle W. van den Maagdenberg, Michael T.M. Emmerich, and Gerard J.P. van Westen. Artificial intelligence in multi-objective drug design. *Current Opinion in Structural Biology*, 79:102537, 2023.

[Monteiro and Reynoso-Meza, 2023] Wellington Rodrigo Monteiro and Gilberto Reynoso-Meza. A multi-objective optimization design to generate surrogate machine learning models in explainable artificial intelligence applications. *EURO Journal on Decision Processes*, 11:100040, 2023.

[Opris *et al.*, 2024] Andre Opris, Duc-Cuong Dang, Frank Neumann, and Dirk Sudholt. Runtime analyses of NSGA-III on many-objective problems. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2024, page 1596–1604. ACM Press, 2024.

[Opris, 2025] Andre Opris. A many objective problem where crossover is provably indispensable. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2025, pages 27108–27116. AAAI Press, 2025.

[Purshouse and Fleming, 2007] Robin C. Purshouse and Peter J. Fleming. On the evolutionary optimization of many conflicting objectives. *IEEE Transactions on Evolutionary Computation*, 11(6):770–784, 2007.

[Ren *et al.*, 2024] Shengjie Ren, Chao Bian, Miqing Li, and Chao Qian. A first running time analysis of the strength pareto evolutionary algorithm 2 (SPEA2). In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVIII, page 295–312. Springer, 2024.

[Sharma and Chahar, 2022] Shubhkirti Sharma and Vijay Chahar. A comprehensive review on multi-objective optimization techniques: Past, present and future. *Archives of Computational Methods in Engineering*, 29:5605–5633, 07 2022.

[Stewart *et al.*, 2008] Theodor Stewart, Oliver Bandte, Heinrich Braun, Nirupam Chakraborti, Matthias Ehrgott, Mathias Göbelt, Yaochu Jin, Hirotaka Nakayama, Silvia Poles, and Danilo Di Stefano. *Real-World Applications of Multi-objective Optimization*, pages 285–327. Springer, 2008.

[Wietheger and Doerr, 2023] Simon Wietheger and Benjamin Doerr. 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, pages 5657–5665. IJCAI Organization, 2023.

[Wietheger and Doerr, 2024] Simon Wietheger and Benjamin Doerr. Near-tight runtime guarantees for many-objective evolutionary algorithms. In *Proceedings of the International Conference on Parallel Problem Solving from Nature*, PPSN XVIII, pages 153–168, 2024.

[Yannibelli *et al.*, 2020] Virginia Yannibelli, Elina Pacini, David Monge, Cristian Mateos, and Guillermo Rodriguez. A comparative analysis of NSGA-II and NSGA-III for autoscaling parameter sweep experiments in the cloud. *Scientific Programming*, 2020(1):4653204, 2020.

[Zapotecas-Martínez *et al.*, 2023] Saúl Zapotecas-Martínez, Carlos A. Coello Coello, Hernán E. Aguirre, and Kiyoshi Tanaka. Challenging test problems for multi- and many-objective optimization. *Swarm and Evolutionary Computation*, 81:101350, 2023.

[Zhang *et al.*, 2021] Huanhuan Zhang, Jigeng Li, Mengna Hong, and Yi Man. Chapter 16 - artificial intelligence algorithm-based multi-objective optimization model of flexible flow shop smart scheduling. In *Applications of Artificial Intelligence in Process Systems Engineering*, pages 447–472. Elsevier, 2021.

[Zheng and Doerr, 2022] Weijie Zheng and Benjamin Doerr. Better approximation guarantees for the NSGA-II by using the current crowding distance. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2022, page 611–619. ACM Press, 2022.

[Zheng and Doerr, 2024a] Weijie Zheng and Benjamin Doerr. Runtime analysis for the NSGA-II: Proving, quantifying, and explaining the inefficiency for many objectives. *IEEE Transactions on Evolutionary Computation*, 28(5):1442–1454, 2024.

[Zheng and Doerr, 2024b] Weijie Zheng and Benjamin Doerr. Runtime analysis of the SMS-EMOA for many-objective optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2024, pages 20874–20882. AAAI Press, 2024.

[Zheng *et al.*, 2022] Weijie Zheng, Yufei Liu, and Benjamin Doerr. 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, pages 10408–10416. AAAI Press, 2022.