The First Theoretical Approximation Guarantees for the Non-Dominated Sorting Genetic Algorithm III (NSGA-III)

Renzhong Deng[1], **Weijie Zheng**[1*], **Benjamin Doerr**[2]
[1] School of Computer Science and Technology, National Key Laboratory of Smart Farm Technologies and Systems, International Research Institute for Artificial Intelligence,
Harbin Institute of Technology, Shenzhen
[2]Laboratoire d'Informatique (LIX), CNRS, École Polytechnique,
Institut Polytechnique de Paris, Palaiseau, France
dengrenzhong@stu.hit.edu.cn, zhengweijie@hit.edu.cn, doerr@lix.polytechnique.fr

Abstract

This work conducts a first theoretical analysis studying how well the NSGA-III approximates the Pareto front when the population size N is less than the Pareto front size. We show that when N is at least the number N_r of reference points, then the approximation quality, measured by the maximum empty interval (MEI) indicator, on the OneMinMax benchmark is such that there is no empty interval longer than $\lceil\frac{(5-2\sqrt{2})n}{N_r-1}\rceil$. This bound is independent of N, which suggests that further increasing the population size does not increase the quality of approximation when N_r is fixed. This is a notable difference to the NSGA-II with sequential survival selection, where increasing the population size improves the quality of the approximations. We also prove two results indicating approximation difficulties when $N < N_r$. These theoretical results suggest that the best setting to approximate the Pareto front is $N_r = N$. In our experiments, we observe that with this setting the NSGA-III computes optimal approximations, very different from the NSGA-II, for which optimal approximations have not been observed so far.

1 Introduction

The non-dominated sorting genetic algorithm II (NSGA-II) [Deb *et al.*, 2002] is the most widely used multi-objective evolutionary algorithm (MOEA). The recent first mathematical runtime analysis of this algorithm [Zheng *et al.*, 2022; Zheng and Doerr, 2023] has inspired many theoretical works on domination-based MOEAs such as the SPEA2 [Zitzler *et al.*, 2001], the SMS-EMOA [Beume *et al.*, 2007], and the NSGA-III [Deb and Jain, 2014]. Notable results include [Bian and Qian, 2022; Doerr and Qu, 2023a; Doerr and Qu, 2023b; Doerr and Qu, 2023c; Dang *et al.*, 2023; Bian *et al.*, 2023; Dinot *et al.*, 2023; Wietheger and Doerr, 2023; Zheng and Doerr, 2024b; Zheng and Doerr, 2024c; Zheng *et al.*, 2024; Opris *et al.*, 2024; Ren *et al.*, 2024; Doerr *et al.*, 2025; Alghouass *et al.*, 2025; Doerr *et al.*, 2025; Li *et al.*, 2025; Opris, 2025a; Opris, 2025b]. Interestingly, these results suggest that the slightly less prominent algorithms SPEA2, SMS-EMOA, and NSGA-III are more powerful than the NSGA-II, at least when the number of objectives is three or more. This is the reason why in this work we concentrate on one of them, namely the NSGA-III.

The mathematical runtime analysis of evolutionary algorithms [Neumann and Witt, 2010; Auger and Doerr, 2011; Jansen, 2013; Zhou *et al.*, 2019; Doerr and Neumann, 2020] so far has almost exclusively regarded the complexity of computing the full Pareto front. In practice, this is often not feasible, because the Pareto front is too large, and it may also not be desirable, since ultimately a human decision maker has to select one of the computed solutions as the solution to be adopted. For this reason, we shall discuss the approximation qualities of the NSGA-III in this work.

In the first and only mathematical work discussing the approximation abilities of one of the above-named algorithms, Zheng and Doerr [2022; 2024a] analyzed how well the NSGA-II approximates the Pareto front of OneMinMax when its population size N is less than the size of the Pareto front. They first observed that the classic NSGA-II, which first computes the crowding distance and then, based on these numbers, selects the next population, can compute very bad approximations, creating empty intervals on the Pareto front by arbitrary factors larger than what an optimal approximation displays. This can be overcome by using the sequential NSGA-II proposed in [Kukkonen and Deb, 2006], which removes individuals sequentially, always updating the crowding distance values after each removal, or the steady-state NSGA-II [Durillo *et al.*, 2009], which generates only one offspring per iteration and hence also removes only one individual per iteration. For these two variants of the NSGA-II, it was proven that within an expected number of $O(Nn \log n)$ function evaluations, approximations with largest empty interval size MEI $\le \max\{\frac{2n}{N-3}, 1\}$ are computed. Since the optimal MEI value for OneMinMax and population size N is $\lceil\frac{n}{N-1}\rceil$, this is essentially a 2-approximation. No such

*Corresponding author.

results exist for any other domination-based algorithms, in particular, not for the NSGA-III.

Our contributions: To fill this research gap, this work will conduct the first analysis of the approximation ability of the NSGA-III, and will detect some notable differences to the NSGA-II. Since ONEMINMAX is the only benchmark for which the approximation abilities of the NSGA-II-type algorithms were studied, in this first approximation work for the NSGA-III we shall also regard this bi-objective problem. We are aware of the fact that generally the NSGA-III is seen as an algorithm for many-objective optimization, but in this first work our focus is on a comparison with the NSGA-II via the results obtained in [Zheng and Doerr, 2024a], so for that reason we restrict ourselves to two objectives. From our proofs, we would conjecture that our findings can be generalized to more objectives.

The main approximation guarantee we prove is that with a population size N at least as the number N_r of reference points, the NSGA-III computes approximations to the Pareto front of ONEMINMAX with MEI $\le \lceil \frac{(5-2\sqrt{2})n}{N_r-1} \rceil$, and this within an expected number of $O(Nn^c \log n)$ function evaluations, where $c = \lceil \frac{2(2-\sqrt{2})}{N_r-1} \rceil$. Recalling that the optimal MEI is MEI $= \lceil \frac{n}{N-1} \rceil$, we see that also the NSGA-III can compute constant factor approximations. Our result shows this factor to be at most $5 - 2\sqrt{2} \approx 2.17$, slightly larger than the factor of 2 shown for the sequential and steady-state NSGA-II.

We also prove that when $N < N_r$, the approximation can be worse than an optimal one by a factor of $\Omega(\log n)$. These results suggest that the number of reference points is best set to be equal to the population size, that is, $N_r = N$. This is consistent with (and thus supports) the suggestion (without theoretical explanation) to take $N_r \approx N$ made in the original NSGA-III paper [Deb and Jain, 2014].

Experiments are conducted to see how the NSGA-III approximates the Pareto front for ONEMINMAX. The results show that with $N_r = N$, the NSGA-III performs better than the NSGA-II with sequential survival selection, and always reaches the optimal approximation. This observation suggests that proving a tighter approximation bound is an interesting target for future research. The experiments also consider the case where $N_r > N$, and show that the approximation ability of the NSGA-III becomes worse as N_r increases, and even the extremal points of the population can be lost (which cannot happen for the NSGA-II as these points have infinite crowding distance). To verify the generalizability of the above theoretical findings and experimental observations on ONEMINMAX, we also conduct experiments for the popular LOTZ benchmark. We again observe that the setting of $N_r = N$ results in optimal approximations, and that large numbers of reference points, that is, $N_r > N$, lead to worse approximations.

In summary, our results show that also the NSGA-III can compute constant-factor approximations of the Pareto front. Different from the NSGA-II (with sequential survival selection or in the steady-state mode), the absolute population size is less important (in particular, increasing the population size does not give better approximations), but the relation to the number of reference points is important for the approximation ability of the NSGA-III. In particular, we observe that for approximating the Pareto front, it appears best that N_r and N are very close, in contrast to the existing result for computing the full Pareto front [Wietheger and Doerr, 2023; Opris *et al.*, 2024], which all require N to be at least a constant factor larger than N_r.

Algorithm 1: NSGA-III

Let the initial population P_0 be composed of N individuals chosen independently and uniformly at random from $\{0,1\}^n$;
for $t = 0, 1, 2, \ldots$ **do**
    Generate the offspring population Q_t with size N;
    Use `fast-non-dominated-sort()` [Deb *et al.*, 2002] to divide $R_t = P_t \cup Q_t$ into $F_1, F_2, \ldots$;
    Find $i^* \ge 1$ such that $\sum_{i=1}^{i^*-1} |F_i| < N$ and $\sum_{i=1}^{i^*} |F_i| \ge N$;
    $Z_t \leftarrow \bigcup_{i=1}^{i^*-1} F_i$;
    Use Algorithm 3 to select $\tilde{F}_{i^*} \subseteq F_{i^*}$ such that $|Z_t \cup \tilde{F}_{i^*}| = N$;
    $P_{t+1} \leftarrow Z_t \cup \tilde{F}_{i^*}$;

2 Preliminaries

This section will give a brief introduction on the NSGA-III, the algorithm to analyze, ONEMINMAX, the benchmark to optimize, and the approximation metric that we will use.

2.1 NSGA-III

The overall framework of the NSGA-III is presented in Algorithm 1. The NSGA-III, a variant of the NSGA-II designed for many objectives, was proposed by Deb and Jain [2014], and its first runtime analysis was conducted by Wietheger and Doerr [2023]. Same as the NSGA-II, the NSGA-III maintains a population P_t of a fixed size N and generates an offspring population Q_t of the same size in each iteration. Also the NSGA-III will remove N individuals from the combined population $R_t = P_t \cup Q_t$ and first uses the non-dominated sorting [Deb *et al.*, 2002] to divide R_t into $F_1, F_2, \ldots$. The NSGA-III only differs in the secondary criterion used in the critical front F_{i^*} for the survival selection. Instead of the crowding distance in the NSGA-II, the NSGA-III uses the following reference point mechanism.

Initially, the objective function values are normalized, and subsequently, the normalized individuals are associated with reference points for selection. We adopted the structured reference point placement method used in [Deb and Jain, 2014], which is based on the systematic approach by [Das and Dennis, 1998]. In this approach, these reference points are uniformly distributed on the normalized hyperplane, specifically on a $(M-1)$-dimensional simplex, where each axis intercept is set to 1. The total number of reference points N_r for

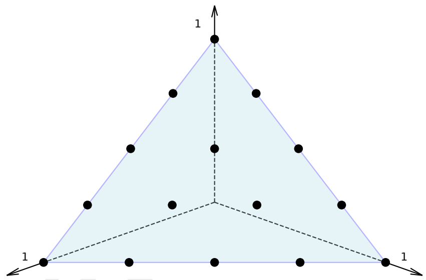

Figure 1: Structured reference points set for a three-objective problem with p = 4 [Deb and Jain, 2014].

an M-objective problem with p divisions along each objective is given by $N_r = \binom{M+p-1}{p}$. Figure 1 shows an example with $M = 3$ and $p = 4$. For bi-objective problems, we have $N_r = p + 1$. Due to the evenly distributed nature of reference points, the ideal scenario is for different individuals in the offspring population P_{t+1} to be associated with different reference points. The greater the number and the more uniform the distribution of reference points associated with individuals in P_{t+1} , the higher its diversity. In the selection of the critical front F_{i^*} of the NSGA-III, the number of individuals already selected to P_{t+1} and associated with a reference point r is referred to as the niche count ρ_r of that reference point. The algorithm prioritizes selecting individuals associated with the reference point that has the smallest niche count ρ if such individuals exist. This strategy increases the number of reference points associated with individuals in P_{t+1}, thereby enhancing the diversity of the population. Algorithms 2 and 3 present the algorithmic frameworks for the normalization and selection processes, respectively. For further details, please refer to [Deb and Jain, 2014; Blank *et al.*, 2019; Wietheger and Doerr, 2023].

2.2 OneMinMax and Approximation Metric

As mentioned before, the bi-objective OneMinMax [Giel and Lehre, 2010] is the only benchmark used in the theoretical community to analyze the approximation ability of the NSGA-II-type algorithms [Zheng and Doerr, 2022; Zheng and Doerr, 2024a]. For a meaningful comparison, this work also chooses this bi-objective benchmark (and also the approximation metric used in these works) to analyze the approximation ability of the NSGA-III. We believe the techniques and the insights obtained in this work will be useful for analyzing problems with many objectives.

The OneMinMax function is defined as follows, and we consider its maximization.

Definition 1 ([Giel and Lehre, 2010]). *For all search points x the objective function $f : \{0,1\}^n \to \mathbb{N} \times \mathbb{N}$ is defined by*

$$f(x) = (f_1(x), f_2(x)) = \left(\sum_{i=1}^{n} x_i, n - \sum_{i=1}^{n} x_i\right).$$

In the language of multi-objective optimization, we say that x *weakly dominates* y, denoted as $x \succeq y$, if $f_1(x) \geq f_1(y)$ and $f_2(x) \geq f_2(y)$. If at least one of the two inequalities is strict, we say that x *dominates* y, denoted as $x \succ y$. If x is not dominated by any individual in $\{0,1\}^n$, it is called *Pareto optimal*, and the set of function values corresponding to all Pareto optimal solutions forms the *Pareto front*. It is easy to see that the Pareto front of OneMinMax is $\{(0,n),(1,n-1),\ldots,(n,0)\}$, with a size of $n+1$.

Algorithm 2: Normalization

Input: $F_1,\ldots,F_{i^*}$: non-dominated fronts;
$f = (f_1,\ldots,f_M)$: objective function;
$z^* \in \mathbb{R}^M$: observed min in each objective;
$z^w \in \mathbb{R}^M$: observed max in each objective;
$E \subseteq \mathbb{R}^M$: extreme points from previous iteration (initially $\{\infty\}^M$);

for $j = 1$ **to** M **do**
  $\hat{z}^*_j = \min\{z^*_j, \min_{z\in Z} f_j(z)\}$;
  Determine an extreme point $e^{(j)}$ in the jth objective from $Z \cup E$ using an achievement scalarization function;
$valid \leftarrow$ **False**;
if $e^{(1)},\ldots,e^{(M)}$ are linearly independent **then**
  $valid \leftarrow$ **True**;
  Let H be the hyperplane spanned by $e^{(1)},\ldots,e^{(M)}$;
  for $j = 1$ **to** M **do**
    $I_j \leftarrow$ the intercept of H with the jth objective axis;
    if $I_j \geq \epsilon_{\text{nad}}$ **and** $I_j \leq z^w_j$ **then**
      $\hat{z}^{\text{nad}}_j \leftarrow I_j$;
    else
      $valid \leftarrow$ **False**;
      break;
if $valid$ =**False then**
  for $j = 1$ **to** M **do**
    $\hat{z}^{\text{nad}}_j = \max_{x\in F_1} f_j(x)$;
for $j = 1$ **to** M **do**
  if $\hat{z}^{\text{nad}}_j < \hat{z}^*_j + \epsilon_{\text{nad}}$ **then**
    $\hat{z}^{\text{nad}}_j = \max_{x\in F_1\cup\cdots\cup F_{i^*}} f_j(x)$;
Define $f^n_j(x) = (f_j(x) - \hat{z}^*_j)/(\hat{z}^{\text{nad}}_j - \hat{z}^*_j), j \in [1..M]$;

As in the NSGA-II's approximation theory [Zheng and Doerr, 2022; Zheng and Doerr, 2024a], we will use the maximum empty interval (MEI) metric (see Definition 2) to evaluate how the algorithm approximates the Pareto front w.r.t. OneMinMax. Note that the MEI can be easily transferred to other commonly used approximation metrics such as ϵ-dominance [Laumanns *et al.*, 2002] or Hypervolume [Zitzler and Thiele, 1998], see [Zheng and Doerr, 2024a].

Definition 2 ([Zheng and Doerr, 2022; Zheng and Doerr, 2024a]). *Let $S = \{(s_1, n-s_1),\ldots,(s_m, n-s_m)\}$ be a subset of Pareto front M of* OneMinMax*. Let $j_1, j_2,\ldots,j_m$ be the sorted list of $s_1,\ldots,s_m$ in the increasing order (ties broken uniformly at random). We define the maximum empty*

Algorithm 3: Selection based on a set U of reference points when maximizing the function f

Input: Z_t: the multi-set of already selected individuals; F_{i^*}: the multi-set of individuals to choose from;

$f_n \leftarrow$ Normalize($f, Z = Z_t \cup F_{i^*}$) using Algorithm 2;
Associate each individual $x \in Z_t \cup F_{i^*}$ to the reference point $\mathrm{rp}(x)$;
For each reference point $r \in U$, let ρ_r denote the number of (already selected) individuals in Z_t associated with r;
$U' \leftarrow U$, $\tilde{F}_{i^*} \leftarrow \emptyset$;
while True **do**
    Let $r_{\min} \in U'$ be such that $\rho_{r_{\min}}$ is minimal (break ties randomly);
    Let $x_{r_{\min}} \in F_{i^*} \setminus \tilde{F}_{i^*}$ be the individual associated with $r_{\min}$ that minimizes the distance between $f^n(x_{r_{\min}})$ and $r_{\min}$ (break ties randomly);
    if $x_{r_{\min}}$ exists **then**
        $\tilde{F}_{i^*} \leftarrow \tilde{F}_{i^*} \cup \{x_{r_{\min}}\}$;
        $\rho_{r_{\min}} \leftarrow \rho_{r_{\min}} + 1$;
        if $|Z_t| + |\tilde{F}_{i^*}| = N$ **then**
            break all and **return** $\tilde{F}_{i^*}$;
    else
        $U' \leftarrow U' \setminus \{r_{\min}\}$;

interval size of S, denoted by $\mathrm{MEI}(S)$, *as*

$$\max\{j_1, n - j_m, j_{i+1} - j_i \mid i = 1, \ldots, m-1\}.$$

For $n \in \mathbb{N}_{\geq 2} := \{i \geq 2 \mid i \in \mathbb{N}\}$, *we further define* $\mathrm{MEI}_{\mathrm{opt}}(N) := \min\{\mathrm{MEI}(S) \mid S \subseteq M, |S| \leq N, (0,n) \in S, (n,0) \in S\}$.

Obviously, this is the smallest MEI *that an MOEA with a fixed population size N can obtain when the extremal points $(0,n)$ and $(n,0)$ are covered.*

The optimal MEI value is shown in the following. Our work will obtain the upper bound of the MEI that the NSGA-III will achieve w.r.t. OneMinMax.

Lemma 3 ([Zheng and Doerr, 2022; Zheng and Doerr, 2024a])**.** *For all $N \in \mathbb{N}_{\geq 2}$, we have* $\mathrm{MEI}_{\mathrm{opt}}(N) = \lceil \frac{n}{N-1} \rceil$.

3 Approximation Guarantee when $N_r \leq N$

In this section, we will theoretically prove that when the number of reference points $N_r \leq N$, the NSGA-III can effectively approximate the Pareto front of OneMinMax.

3.1 Reach Extremal Points

To ease the discussion, here we call any optimal solution of one objective in a multi-objective problem an *extremal point*. Note that it is not the same as the extreme point/vector used to determine the hyperplane, see Algorithm 2. For OneMinMax, 0^n and 1^n are the only extremal points. We will consider whether the two extremal points 0^n and 1^n can be reached.

To achieve this, the following lemma will first show that for $N_r \leq N$, the maximal and minimal objective values in the combined population survives. The key point is that after the normalization, such values will be mapped to $(1,0)$ or $(0,1)$. Note that $(1,0)$ and $(0,1)$ are reference points. Hence, the above values is closest to such reference points. Since $N_r \leq N$ and all solutions are in F_1, we know that to select N individuals in the survival selection in Algorithm 3, the final $\rho_{r_{\min}}$ must be greater than or equal to 1. Thus, before $\rho_{r_{\min}} > 1$, at least one individual with the maximal or minimal objective value (normalized to $(1,0)$ or $(0,1)$) will be selected into the next generation. Due to the page limit, all proofs are included in the full version [Deng *et al.*, 2025].

Lemma 4. *Consider using the NSGA-III with population size N to optimize* OneMinMax *with problem size n and suppose that $\epsilon_{\mathrm{nad}} \geq n$. Let $N < n+1$ and $N_r \leq N$. Let $z_1^{\min} := \min\{f_1(x) \mid x \in R_t\}$ and $z_1^{\max} := \max\{f_1(x) \mid x \in R_t\}$, where $R_t = P_t \cup Q_t$ is the combined parent and offspring population. Then P_{t+1} will contain two individuals x, y such that $f_1(x) = z_1^{\min}$ and $f_1(y) = z_1^{\max}$.*

Note that the offspring generation operator (see Step 3 in Algorithm 1) will not increase $z_1^{\min}$ or decrease $z_1^{\max}$. Then Lemma 4 shows that the minimal value of any objective will not increase and the maximal value of any objective will not decrease for all generations. Hence, the extremal points 0^n and 1^n will not be removed once they are reached. Thus, we focus on the time to decrease the minimal value of f_1 to 0 to reach 0^n, and the time to increase the maximal value of f_1 to n to reach 1^n. Lemma 5 provides an upper bound on the runtime for the population to cover these two extremal points.

Lemma 5. *Consider using the NSGA-III with population size N to optimize* OneMinMax *with problem size n and suppose that $\epsilon_{\mathrm{nad}} \geq n$. Let $N < n+1$ and $N_r \leq N$. Then within an expected number of $O(n \log n)$ iterations, that is, an expected number of $O(Nn \log n)$ fitness evaluations, the two extremal points 0^n and 1^n will be reached for the first time, and both points will be kept in all future populations.*

3.2 Good Approximation Guarantee

Lemma 5 shows that the population will always contain 0^n and 1^n once they are reached for the first time. From basic calculations based on Algorithm 2, we obtain the following clear form for the normalized function value when the population contains both 0^n and 1^n.

Lemma 6. *Consider using the NSGA-III with population size N to optimize* OneMinMax *with problem size n and suppose that $\epsilon_{\mathrm{nad}} \geq n$. Let $N < n+1$ and $N_r \leq N$. Assume that the two extremal points 0^n and 1^n are included in the population. Then for any individual x in the population, its function value $f(x)$ is normalized to $f^n(x) = \frac{1}{n} f(x)$.*

With the clear form of the normalized function value obtained in Lemma 6, we easily map the function values to the normalized space so that we obtain the upper bound of the distance between the normalized function value and the associated reference point in Lemma 7 below.

Lemma 7. *Consider using the NSGA-III with population size N to optimize* OneMinMax *with problem size n and suppose that $\epsilon_{\mathrm{nad}} \geq n$. Let $N < n+1$, $N_r \leq N$, and let* *individual x be associated with reference point $r = (r_1, r_2)$. Assume that the two extremal points 0^n and 1^n are included in the population. Normalize $f(x)$ to $f^n(x) = (f_1^n(x), f_2^n(x))$. Then $f^n(x)$ is located in the non-negative region of the reference point plane and $|f_1^n(x) - r_1| \leq \frac{2-\sqrt{2}}{N_r-1}$.*

Since 0^n and 1^n are included in the population, Lemma 7 indicates a good mapping between the evenly distributed reference points N_r and the desired good approximation of the evenly distributed function values. It will be our key proof idea for our approximation guarantee in Theorem 11.

For better description, we call a reference point *active* or *activated* if it has at least one associated individual in the combined parent and offspring population before the survival selection. Now we first show in Lemma 8 that once a reference point is active, it will remain active forever. The key to the proof is that in Algorithm 3, each reference point will be selected at least once. Hence, at least one individual associated with the active reference point is retained.

Lemma 8. *Consider using the NSGA-III with population size N to optimize* ONEMINMAX *with problem size n and suppose that $\epsilon_{\text{nad}} \geq n$. Let $N < n + 1$ and $N_r \leq N$. Assume that the two extremal points 0^n and 1^n are included in the population. Then, if a reference point r has individuals associated with it, there will always be at least one individual associated with it in future generations.*

As active reference points will remain active, we now calculate the time to activate all reference points, that is, to reach a status when all reference points have their associated individuals. From Lemma 7 we have the estimate between the active reference point and its associated individual's normalized function value. Then we use the waiting time argument to estimate the time from an active reference point to generate a neighbor non-activated reference point. Hence, the overall time for activating all reference points is obtained in Lemma 9.

Lemma 9. *Consider using the NSGA-III with population size N to optimize* ONEMINMAX *with problem size n and suppose that $\epsilon_{\text{nad}} \geq n$. Let $N < n + 1$ and $N_r \leq N$. Let $c := \lceil \frac{2(2-\sqrt{2})n}{N_r-1} \rceil$. Assume that the two extremal points 0^n and 1^n are included in the population. Then after an expected number of $O(n^c \log n)$ iterations, that is, an expected number of $O(Nn^c \log n)$ fitness evaluations, each reference point r has at least one individual x associated with it and this property will be kept for all future time.*

Although Lemma 9 estimates the runtime for activating all reference points, the individuals associated with the same reference point do not necessarily share the same normalized function value. Hence, we now use Lemma 7 to estimate the distance between individuals associated to neighboring reference points, and then obtain an upper bound on MEI (see the following lemma).

Lemma 10. *Consider using the NSGA-III with population size N to optimize* ONEMINMAX *with problem size n and suppose that $\epsilon_{\text{nad}} \geq n$. Let $N < n + 1$, $N_r \leq N$, and let t_0 be the first generation such that the two extremal points $0^n$and 1^n are included in the population. Assume that each reference point r has at least one individual x associated with it. Then for any $t \geq t_0$, we have* $\text{MEI} \leq \lceil \frac{(5-2\sqrt{2})n}{N_r-1} \rceil$.

Hence, from Lemmas 5, 9 and 10, we obtain the following theorem regarding the approximation ability of NSGA-III.

Theorem 11. *Consider using the NSGA-III with population size N to optimize* ONEMINMAX *with problem size n and suppose that $\epsilon_{\text{nad}} \geq n$. Let $N < n+1$ and $N_r \leq N$. Let $c := \lceil \frac{2(2-\sqrt{2})n}{N_r-1} \rceil$. Then after an expected number of $O(Nn^c \log n)$ fitness evaluations, a population containing the two extremal points 0^n and 1^n and with* $\text{MEI} \leq \lceil \frac{(5-2\sqrt{2})n}{N_r-1} \rceil$ *is reached and both properties will be kept for all future time.*

From Theorem 11, we see both the upper bound of the MEI value and the runtime to reach such upper bound heavily depend on N_r, the number of reference points. It is quite different from the NSGA-II (with sequential survival selection) whose upper bound $\text{MEI} \leq \max\{\frac{2n}{N-3}, 1\}$ [Zheng and Doerr, 2024a] merely relies on N, the population size. Note that the optimal $\text{MEI} = \lceil \frac{n}{N-1} \rceil$ merely depends on N, the number of points, as well. Besides, Theorem 11 shows that the minimal MEI upper bound of $\lceil \frac{(5-2\sqrt{2})n}{N-1} \rceil$ that the NSGA-III can reach happens when $N_r = N$. Although this upper bound is slightly worse than the one for the NSGA-II with sequential survival selection discussed before, our experiments in Section 5 shows the NSGA-III always reaches the optimal MEI, and performs better than the NSGA-II.

4 Possible Difficulty for Large $N_r > N$

In contrast to the approximation guarantee in the previous section, this section will show that NSGA-III may lose extremal points and yield poor approximation quality when N_r is sufficiently large. It is quite different from the existing theoretical results [Wietheger and Doerr, 2023; Opris *et al.*, 2024] that require the large enough N_r for their analyses.

4.1 NSGA-III Can Lose Extremal Points

Recall the definition of the extremal pointed in the above section. For ONEMINMAX, 0^n and 1^n are the only extremal points. Being the optima of single objectives, extremal points are particularly important and should therefore not be removed from the population. In the NSGA-II, for the multiset of any extremal point, at least one of the repetitions will have infinite crowding distance and will survive to the next generation if the population size is large enough, say at least 3 for ONEMINMAX. However, the NSGA-III has no such a special treatment to ensure the extremal points survive to the next generation. Instead, the removal is determined by the numbers of already chosen individuals associated to the reference points, and the extremal points have the chance to be removed. The following lemma gives an example where the extremal point can be removed.

Lemma 12. *Consider using the NSGA-III with population size N to optimize* ONEMINMAX *with problem size n and suppose that $\epsilon_{\text{nad}} \geq n$. Let $N \leq (n + 1)/2$ and $N_r \geq 2N$. Assume that before the survival selection, in the combined parent and offspring population R_t that contains 0^n and 1^n,* *each reference point has at most one associated individual. Then after the survival selection, the probabilities to remove one specific extremal point, to remove at least one specific extremal point, and to remove both extremal points are* $\frac{1}{2}$, $\frac{3}{4}+\frac{1}{4(2N-1)}$, *and* $\frac{1}{4}-\frac{1}{4(2N-1)}$ *respectively.*

The above lemma shows that with a good probability, the survival selection will lose extremal points for this example.

4.2 Possible Bad Approximation Quality

In addition to the above example that the extremal points can be removed, here we discuss an example where the survival selection of the NSGA-III can lead to a quite bad approximation from a very promising situation. This situation is essentially similar to the one used to show the possible difficulty of the traditional NSGA-II in [Zheng and Doerr, 2024a].

Lemma 13. *Consider using the NSGA-III with problem size* N *to optimize* ONEMINMAX *with problem size* n *and suppose that* $\epsilon_{\text{nad}} \geq n$. *Let* n *be odd and* $N=(n+1)/2$ *and* $N_r=2N$. *Assume that the combined parent and offspring population* R_t *covers the full Pareto front. Then the next population has the expected* MEI *value of* $\Omega(\log n)$.

The optimal MEI value of $N=(n+1)/2$ is $\lceil\frac{n}{N-1}\rceil \leq 4$. The above lemma shows that even from the above optimal-looking combined population R_t, the MEI value becomes $\Omega(\log n)$ in expectation. Due to the complicated process of the NSGA-III, we currently do not know how often the above situation happens. However, our experiments in Section 5 show that the difficulties of the NSGA-III with large number of reference points exist indeed.

As mentioned before, this work focuses on the approximation ability of the original NSGA-III. Hence, here we will not discuss the strategies to overcome the above difficulty stemming from losing extremal points, but leave it as interesting future research.

5 Experiments

Previous sections gave theoretical approximation results for the NSGA-III on ONEMINMAX. This section conducts extensive experiments to address the following three research questions.

(1) The impact of the number of reference points N_r. Our theorem shows that the best approximation ability of the NSGA-III on ONEMINMAX happens when $N_r=N$. We will see whether it is experimentally true for ONEMINMAX. Besides, Section 4 proved the difficulties of the NSGA-III with $N_r>N$ on special cases. It remains unknown whether it actually happens.

(2) The actual approximation ability. The upper bound of the MEI approximation metric obtained in Theorem 11 for ONEMINMAX is a factor of $5-2\sqrt{2}$ larger than the optimal value, and is slightly worse than the theoretical bound for the NSGA-II (with sequential survival selection). We will see how well the NSGA-III experimentally approximates, together with its comparison with the NSGA-II.

(3) The verification of the above findings beyond ONEMINMAX. The above two concerns are for the verification on ONEMINMAX as in our theoretical results. A natural question is about the behaviors of the NSGA-III on other problems.

Generations	[1..100]	[3001..3100]
NSGA-II’	(11,12,12)	(11,12,12)
$N_r=\lceil N/4\rceil$	(33,33,33)	(33,33,33.25)
$N_r=\lceil N/2\rceil$	(17,17,17)	(17,17,17)
$N_r=N$	(9,9,9)	(9,9,9)
$N_r=2N$	(187,194,203)	(186,194,203)
$N_r=4N$	(229,237,244)	(236,240,245)
$N_r=8N$	(228,234,240)	(229,240,248)

Table 1: The 1st, 2nd, and 3rd quartiles (displayed in the format of $(\cdot,\cdot,\cdot)$) for the MEI within 100 generations and 20 independent runs. Generations $[1..100]$ and $[3001..3100]$ after T_{start} on ONEMINMAX with $n=601$ and $N=76$ are regarded separately. Here, T_{start} is the generation number when the population contains both extremal points for the first time. For the case of losing extremal points for $N_r>N$, it is set to be $T_{\max}$, the maximal value of T_{start} reported for the algorithm with the settings $N_r \leq N$ in 20 independent runs. Note that $\text{MEI}_{\text{opt}}=\lceil n/(N-1)\rceil=9$.

5.1 Experimental Settings

To address the first two concerns w.r.t. ONEMINMAX, we adopt the same settings in the only theoretical approximation works of the NSGA-II [Zheng and Doerr, 2022; Zheng and Doerr, 2024a]. That is, we consider the problem size $n=601$ and population sizes $N=301,151,76$ ($N=\lceil(n+1)/2\rceil,\lceil(n+1)/4\rceil,\lceil(n+1)/8\rceil$) in these works for the easy comparison to the NSGA-II (with sequential survival selection, denoted as NSGA-II’ in the following). We set the number of reference points $N_r=\lceil N/4\rceil,\lceil N/2\rceil,N,2N,4N,8N$, with the first three for $N_r\leq N$ discussed in Section 3 and the last three for $N_r>N$ discussed in Section 4.

To tackle the last concern, we experimentally consider another popular benchmark, LOTZ, and the optimal MEI $=\lceil\frac{n}{N-1}\rceil$. We set the problem size $n=120$, the setting used in [Zheng and Doerr, 2023], and the population size $N=\lceil(n+1)/2\rceil,\lceil(n+1)/4\rceil,\lceil(n+1)/8\rceil$ and the number of reference points $N_r=\lceil N/4\rceil,\lceil N/2\rceil,N,2N,4N,8N$, same settings as for ONEMINMAX.

20 independent runs are conducted. We regard these numbers are enough as the collected data are quite concentrated.

5.2 Experimental Results

Due to the limited space, here we only report the results for $N=\lceil(n+1)/8\rceil$ (that is, $N=76$ for ONEMINMAX and $N=16$ for the LOTZ), while the results for other population sizes are included in the full version [Deng *et al.*, 2025]. Note that experimental findings obtained for this setting of N also similarly hold for other N.

The Impact of N_r

From the stable MEI values reported in Table 1 for 20 runs and Figure 2 for one exemplary run, we see that $N_r=N$ achieves the best MEI value, compared with other settings of N_r. We also see that small values of $N_r<N$ leads to significantly better approximation quality against large values of

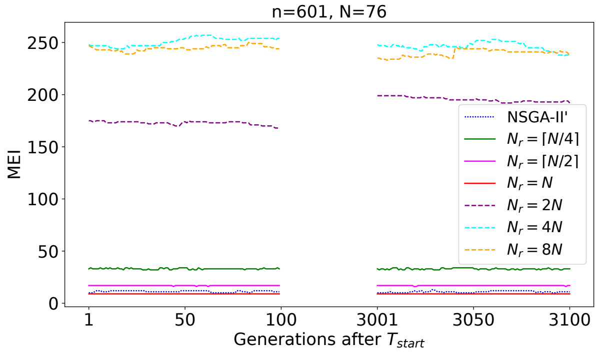

Figure 2: The MEI for generations $[1..100]$ and $[3001..3100]$ after T_{start}, in one exemplary run on ONEMINMAX with $n = 601$ and $N = 76$.

$N_r > N$. Furthermore, in the experiments, we observed that the algorithm terminates only when $T_{\max}$ is reached, which shows that the difficulty to have both 0^n and 1^n exists and verifies the possible difficulty of the NSGA-III with large enough number of reference points discussed in Section 4.

The Actual Approximation Ability

Note that for $n = 601$ and $N = 76$ reported in Table 1, the optimal MEI $= \lceil \frac{n}{N-1} \rceil = 9$. Hence, for the best setting of $N_r = N$, we see all three quartiles reach this optimal value. Together with one exemplary one plotted in Figure 2 and all data in generations $[3001..3100]$), we know that the NSGA-III will reach the optimal MEI for ONEMINMAX, while the NSGA-II' (the sequential version of the NSGA-II analyzed in [Zheng and Doerr, 2024a]) cannot. This result shows that the upper bound $\lceil \frac{(5-2\sqrt{2})n}{N-1} \rceil$ (with $N_r = N$) obtained in Theorem 11 is not tight, and this is left as an interesting question for future research.

The Approximation Ability of the NSGA-III on LOTZ

Analogous to Table 1 and Figure 2 for ONEMINMAX, Table 2 (20 runs) and Figure 3 (one exemplary run) are for the LOTZ problem. We see that $N_r = N$ is the best setting for the approximation ability of the NSGA-III on LOTZ, and this best setting results in the optimal MEI value of 8 (note that the Pareto front for LOTZ is $\{(0, n), (1, n-1), \ldots, (n, 0)\}$, which is the same for ONEMINMAX. Hence, the optimal MEI value is MEI $= \lceil \frac{n}{N-1} \rceil = 8$ for $n = 120$ and $N = 16$).

6 Conclusion

This paper conducted the first theoretical analysis for the approximation performance of the NSGA-III. We showed that the number of reference points plays an essential role for the approximation quality of the NSGA-III, considering the ONEMINMAX benchmark.

In detail, we proved that when the population size N is smaller than the size of the Pareto front, the NSGA-III can achieve an MEI approximation value of $\lceil \frac{(5-2\sqrt{2})n}{N_r-1} \rceil$ on ONEMINMAX within an expected $O(Nn^c \log n)$ function evaluations, where $N_r \le N$ is the number of reference points and $c = \lceil \frac{2(2-\sqrt{2})n}{N_r-1} \rceil$. Then the best upper bound of the approximation quality (only a constant factor of $5 - 2\sqrt{2}$ larger than the optimal MEI) is witnessed for $N_r = N$. However, when $N_r > N$, we proved the possible approximation difficulties of the NSGA-III for some examples.

Generations	After T'_{start}
$N_r = \lceil N/4 \rceil$	(40,40,40)
$N_r = \lceil N/2 \rceil$	(18,18,18)
$N_r = N$	(8,8,8)
$N_r = 2N$	(19,22,24)
$N_r = 4N$	(41,50,61)
$N_r = 8N$	(45,52,63)

Table 2: The 1st, 2nd, and 3rd quartiles (displayed in the format of $(\cdot,\cdot,\cdot)$) for the MEI within 1000 generations after T'_{start} in 20 independent runs on LOTZ with $n = 120$ and $N = 16$. Here, T'_{start} is the generation number when the population contains both extremal points and has MEI $= \lceil \frac{n}{N_r-1} \rceil$ for the first time. For the case of losing extremal points for $N_r > N$, it is set to be $T'_{\max}$, the maximal value of T'_{start} reported for the algorithm with the settings $N_r \le N$ in 20 independent runs. Note that $\text{MEI}_{\text{opt}} = \lceil n/(N-1) \rceil = 8$.

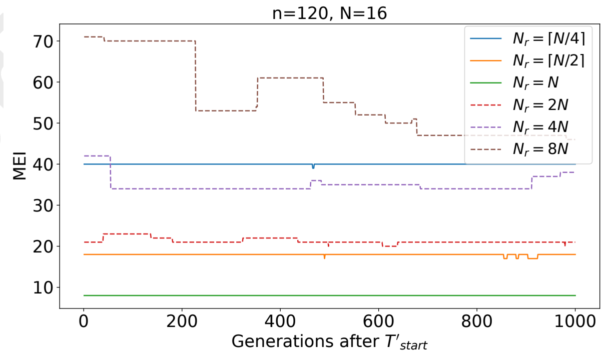

Figure 3: The MEI for generations $[1..1000]$ after T'_{start}, in one exemplary run on the LOTZ problem with $n = 120$ and $N = 16$.

Our experiments verified the above findings, say the good approximation ability of the NSGA-III with $N_r \le N$ and its bad approximation ability for $N_r > N$. Our experiments also showed that the NSGA-III with $N_r = N$ can reach the optimal MEI for approximation, which will guide us to derive a tighter approximation bound as our future interesting research topic.

We note the differences compared to the existing theoretical works. For the approximation ability of the NSGA-II [Zheng and Doerr, 2024a], the population size influences the upper bound of the MEI value. However, there the number of reference points has the essential influence. Besides, different from the existing theory works [Wietheger and Doerr, 2023; Opris *et al.*, 2024], a large enough number of the reference points are required to establish the runtime theory for the full coverage of the Pareto front. This work shows that a large number of reference points can be harmful for the approximation.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No. 62306086, 62350710797), Science, Technology and Innovation Commission of Shenzhen Municipality (Grant No. GXWD20220818191018001), and Guangdong Basic and Applied Basic Research Foundation (Grant No. 2025A1515011936). This research benefited from the support of the FMJH Program PGMO. This work has profited considerably from discussions at the Dagstuhl Seminars 23361 *Multiobjective Optimization on a Budget* and 24271 *Theory of Randomized Optimization Heuristics*.

References

[Alghouass *et al.*, 2025] Yasser Alghouass, Benjamin Doerr, Martin S. Krejca, and Mohammed Lagmah. Proven approximation guarantees in multi-objective optimization: SPEA2 beats NSGA-II. In *International Joint Conference on Artificial Intelligence, IJCAI 2025*. ijcai.org, 2025. Accepted for publication.

[Auger and Doerr, 2011] Anne Auger and Benjamin Doerr, editors. *Theory of Randomized Search Heuristics*. World Scientific Publishing, 2011.

[Beume *et al.*, 2007] Nicola Beume, Boris Naujoks, and Michael Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181:1653–1669, 2007.

[Bian and Qian, 2022] Chao Bian and Chao Qian. Better running time of the non-dominated sorting genetic algorithm II (NSGA-II) by using stochastic tournament selection. In *Parallel Problem Solving From Nature, PPSN 2022*, pages 428–441. Springer, 2022.

[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 *International Joint Conference on Artificial Intelligence, IJCAI 2023*, pages 5513–5521. ijcai.org, 2023.

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

[Dang *et al.*, 2023] 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 *Conference on Artificial Intelligence, AAAI 2023*, pages 12390–12398. AAAI Press, 2023.

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

[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*, 18:577–601, 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*, 6:182–197, 2002.

[Deng *et al.*, 2025] Renzhong Deng, Weijie Zheng, and Benjamin Doerr. The first theoretical approximation guarantees for the non-dominated sorting genetic algorithm III (NSGA-III). *CoRR*, abs/2504.21552, 2025.

[Dinot *et al.*, 2023] Matthieu Dinot, Benjamin Doerr, Ulysse Hennebelle, and Sebastian Will. Runtime analyses of multi-objective evolutionary algorithms in the presence of noise. In *International Joint Conference on Artificial Intelligence, IJCAI 2023*, pages 5549–5557. ijcai.org, 2023.

[Doerr and Neumann, 2020] Benjamin Doerr and Frank Neumann, editors. *Theory of Evolutionary Computation—Recent Developments in Discrete Optimization*. Springer, 2020. Also available at http://www.lix.polytechnique.fr/Labo/Benjamin.Doerr/doerr_neumann_book.html.

[Doerr and Qu, 2023a] Benjamin Doerr and Zhongdi Qu. A first runtime analysis of the NSGA-II on a multimodal problem. *IEEE Transactions on Evolutionary Computation*, 27:1288–1297, 2023.

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

[Doerr and Qu, 2023c] 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 *Conference on Artificial Intelligence, AAAI 2025*, pages 26964–26972. AAAI Press, 2025.

[Durillo *et al.*, 2009] Juan J. Durillo, Antonio J. Nebro, Francisco Luna, and Enrique Alba. On the effect of the steady-state selection scheme in multi-objective genetic algorithms. In *Evolutionary Multi-Criterion Optimization*, pages 183–197. Springer, 2009.

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

[Jansen, 2013] Thomas Jansen. *Analyzing Evolutionary Algorithms – The Computer Science Perspective*. Springer, 2013.

[Kukkonen and Deb, 2006] Saku Kukkonen and Kalyanmoy Deb. Improved pruning of non-dominated solutions based on crowding distance for bi-objective optimization problems. In *Conference on Evolutionary Computation, CEC 2006*, pages 1179–1186. IEEE, 2006.

[Laumanns *et al.*, 2002] Marco Laumanns, Lothar Thiele, Kalyanmoy Deb, and Eckart Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10:263–282, 2002.

[Li *et al.*, 2025] Mingfeng Li, Weijie Zheng, and Benjamin Doerr. Scalable speed-ups for the SMS-EMOA from a simple aging strategy. In *International Joint Conference on Artificial Intelligence, IJCAI 2025*. ijcai.org, 2025. Accepted for publication.

[Neumann and Witt, 2010] Frank Neumann and Carsten Witt. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer, 2010.

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

[Opris, 2025a] Andre Opris. A first runtime analysis of NSGA-III on a many-objective multimodal problem: Provable exponential speedup via stochastic population update. In *International Joint Conference on Artificial Intelligence, IJCAI 2025*. ijcai.org, 2025. Accepted for publication.

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

[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 *Parallel Problem Solving from Nature, PPSN 2024, Part III*, pages 295–312. Springer, 2024.

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

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

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

[Zheng and Doerr, 2024a] Weijie Zheng and Benjamin Doerr. Approximation guarantees for the non-dominated sorting genetic algorithm II (NSGA-II). *IEEE Transactions on Evolutionary Computation*, 2024. In press, https://doi.org/10.1109/TEVC.2024.3402996.

[Zheng and Doerr, 2024b] 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*, 28:1442–1454, 2024.

[Zheng and Doerr, 2024c] Weijie Zheng and Benjamin Doerr. Runtime analysis of the SMS-EMOA for many-objective optimization. In *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 *Conference on Artificial Intelligence, AAAI 2022*, pages 10408–10416. AAAI Press, 2022.

[Zheng *et al.*, 2024] Weijie Zheng, Mingfeng Li, Renzhong Deng, and Benjamin Doerr. How to use the Metropolis algorithm for multi-objective optimization? In *Conference on Artificial Intelligence, AAAI 2024*, pages 20883–20891. AAAI Press, 2024.

[Zhou *et al.*, 2019] Zhi-Hua Zhou, Yang Yu, and Chao Qian. *Evolutionary Learning: Advances in Theories and Algorithms*. Springer, 2019.

[Zitzler and Thiele, 1998] Eckart Zitzler and Lothar Thiele. Multiobjective optimization using evolutionary algorithms—a comparative case study. In *Parallel Problem Solving from Nature, PPSN 1998*, pages 292–301. Springer, 1998.

[Zitzler *et al.*, 2001] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. SPEA2: Improving the strength Pareto evolutionary algorithm. *TIK report*, 103, 2001.