PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

On the Impact of Crossover in Many-Objective Optimization: A Runtime Analysis of NSGA-III

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

Abstract

In recent years, a theoretical understanding has rapidly advanced regarding how popular multi-objective evolutionary algorithms (MOEAs) can optimize many-objective problems. However, the benefits of using crossover in many-objective optimization are theoretically not understood, except for specifically designed benchmark functions tuned to particular crossover operators, and still lag significantly behind its practical use. In this paper, we build upon this line of research and present a theoretical runtime analysis of the widely used NSGA-III algorithm on the classical m-objective m-ONEJUMPZEROJUMP function (m-OJZJ for short). Our results demonstrate that NSGA-III with crossover optimizes m-OJZJ asymptotically faster than NSGA-III without crossover for any number m of objectives for huge parameter regimes. We complement our analysis by providing a lower runtime bound on 4-OJZJ when crossover is turned off.

1 Introduction

Multi-objective evolutionary algorithms (MOEAs), such as the non-dominated sorting genetic algorithm II (NSGA-II) [Deb *et al.*, 2002] and its extension for many-objective optimization, NSGA-III [Deb and Jain, 2014], have been used in thousands of applications and together have received more than 60,000 citations. Applications span a wide range of domains, including machine learning [Bi-wu Zhu *et al.*, 2025], bioengineering [Rashmi *et al.*, 2025], and artificial intelligence [Luukkonen *et al.*, 2023], where also many studies involve four or more objectives. The most prominent EMOA for optimizing bi-objective problems is NSGA-II [Deb *et al.*, 2002] (see [Vijai and P., 2025] for empirical results, or [Zheng *et al.*, 2022; Dang *et al.*, 2023] for rigorous ones), however, which is not well suited for solving problems where the number of objective increases [Campos Ciro *et al.*, 2016; Doerr *et al.*, 2025]. The reason is, that in the bi-objective case, an ordering with respect to the first objective of non-dominated solutions implies also an ordering with respect to the second, which makes the crowding distance, the second tie breaker of NSGA-II, effective. However, this relation breaks down already for three objectives, which indicates the inefficiency of NSGA-II in the many-objective setting. However, NSGA-III, a refinement of NSGA-II, uses reference points instead of crowding distance to ensure that the solution set will be well-distributed across the objective space in a very natural way (see [Deb and Jain, 2014]). It has been shown both theoretically [Wietheger and Doerr, 2023; Opris *et al.*, 2024a] and empirically [Campos Ciro *et al.*, 2016] that NSGA-III effectively optimizes many-objective problems, which is a challenging task, as the size of the Pareto front and the number of incomparable solutions grow exponentially with the number of objectives. However, these first rigorous runtime analyses of the state-of-the-art NSGA-III were published just a few years ago. As a result, its theoretical understanding still lags behind its practical achievements, although some progress has been made since then [Opris, 2025a; Opris, 2026a; Opris, 2026b]. There is still a significant gap in our theoretical understanding of how crossover contributes to many-objective optimization, despite its importance as a fundamental operator in evolutionary computation [Pavai and Geetha, 2016] which hinder our general understanding when and why MOEAs perform well. Beyond up to two papers [Opris, 2025c; Opris, 2026a], which theoretically demonstrate that crossover yields an exponential runtime speedup on handcrafted benchmark functions, we are not aware of any further theoretical results, in contrast to the bi-objective setting [Dang *et al.*, 2023; Doerr and Qu, 2023a]. But further empirical results on many different multi-objective constrained, and unconstrained problems [Sharma *et al.*, 2021], and on multi- and many-objective Knapsack problems [Ishibuchi *et al.*, 2014] show the huge potential of crossover operators in many-objective optimization.

Our contribution: In this paper, we build on the considerations from [Opris, 2025a; Doerr and Qu, 2023a], and provide a theoretical runtime analysis of NSGA-III on the many-objective m-OJZJ$_k$ benchmark with and without crossover for any number of objectives m, and show that NSGA-III without crossover needs $O(m^2 n \ln(n/m) + \delta m n^k (1 + 2n/m)^{m/2}/\mu)$ generations with probability $1 - o(1) - e^{-(\delta-3)m}$ to cover the whole Pareto front. Here, μ denotes the population size, n the problem size, k is a parameter specific to the problem at hand, and $\delta > 3$ is an additional parameter. The expected number of generations is $O(m^2 n \ln(n/m) + m n^k (1 + 2n/m)^{m/2}/\mu)$. One sees that this number is asymptotically the same for a wide range of population sizes μ. This robustness stems from the observation that NSGA-III can retain many individuals with the same fitness vector due to how it associates solutions with and iterates over reference points. This aspect was not considered in [Wietheger and Doerr, 2024]. In particular, our analysis improves the runtime bound given there for population sizes asymptotically larger than the size of the Pareto front, and it even extends the analysis from [Opris, 2025a] to an arbitrary number of objectives. With uniform crossover, we show that with probability $1-o(1)-e^{-(\delta-3)m}$, the number of generations until the whole Pareto front is covered is at most $O(k\delta m^{k+1}n^k c^k\mu/k!)$ for NSGA-III, and the expected number of generations is $O(km^{k+1}n^k c^k\mu/k!)$ for a suitable constant $c>0$. Thus, one obtains a speedup of order $\Omega((k-1)!(1+2n/m)^{m/2}/(\mu^2 c^k m^k))$ in the expected runtime compared to the case without crossover. If $k=\Omega(n/\ln(n))$, $m=O(\log(n))$ and $\mu=O((1+2n/m)^{m/2})$, this speedup even becomes exponential. In a nutshell, our proof applies the arguments from [Doerr and Qu, 2023b] for the bi-objective case sequentially to all blocks. A key difference is that we also need high-probability guarantees for finding a single Pareto-optimal solution in order to obtain all of them in reasonable time in parallel. Finally, we complement our analysis by also providing a lower runtime bound of NSGA-III without crossover on 4-OJZJ of $\Omega(n^k/\mu)$ generations, which is by a factor of $(k-1)!/(\mu^2 d^k)$ larger than the upper runtime bound derived with crossover for a constant $d>0$. This factor is also exponential if $k=\Omega(n/\ln(n))$ and $\mu=\text{poly}(n)$. Extending this lower bound to a larger number of objectives appears considerably more difficult, since the interactions between single objectives becomes more complex, and requires a much deeper understanding of the underlying population dynamics which extends the scope of this paper. However, we expect our results to extend similarly to other MOEAs such as GSEMO, SPEA2, SMS-EMOA, and variants of PAES-25 [Opris, 2025b].

Related work: In single-objective optimization, the benefits of crossover are much better understood than in the many-objective setting. On pseudo-Boolean benchmark problems such as JUMP_k, where a fitness valley of size k must be crossed, it has been rigorously shown that *uniform* crossover yields a speedup depending on k and the crossover probability [Jansen and Wegener, 2002; Kötzing *et al.*, 2011; Dang *et al.*, 2017; Opris *et al.*, 2024b; Opris and Antipov, 2026]. An exponential performance gap in the runtime was proven in [Jansen and Wegener, 2005] on a function REALROYALROAD, which is specifically designed for 1-point crossover. These insights have been used to prove also advantages through crossover for combinatorial optimization problems like shortest paths [Doerr *et al.*, 2012; Doerr *et al.*, 2013], solving complex data clustering problems [Gaeuman and Sutton, 2025] or NP-hard graph problems [Sutton and Lee, 2024], or even in more complex search spaces like permutation spaces [Opris *et al.*, 2025].

In multiobjective optimization, only a few variants of the global simple evolutionary multiobjective optimizer (GSEMO) with crossover have been studied [Qian *et al.*, 2013; Qian *et al.*, 2020; Doerr *et al.*, 2022], and rigorous analyses of NSGA-II with crossover on classical benchmark problems [Doerr and Qu, 2023b] and multi-objective variants of REALROYALROAD [Dang *et al.*, 2023] have been conducted. However, these results were only restricted on bi-objective problems. The theoretical analysis of NSGA-III only succeeded recently. Based on a rigorous analysis of GSEMO on classical benchmark functions [Laumanns *et al.*, 2004], in [Wietheger and Doerr, 2023] the first runtime analysis of NSGA-III on the easy 3-ONEMINMAX problem was conducted. This was then generalized by [Opris *et al.*, 2024a; Opris, 2025a] on more than three objectives, where also key structural insights into the working principles of NSGA-III are given. Also, other pseudo-Boolean functions, where one has also to reach the Pareto front at a first glance, have been analyzed there. Then, in [Wietheger and Doerr, 2024; Opris, 2025a] m-OJZJ has been analyzed, but without investigating crossover. Similar analyses have then also been conducted on other popular MOEAs like the SPEA-2 [Ren *et al.*, 2024] and the SMS-EMOA [Zheng and Doerr, 2024]. First theoretical results which showcase that MOEAs, particularly NSGA-III and SPEA-2, are quite robust with respect to the chosen population size can be found in [Opris, 2026b; Doerr *et al.*, 2026]. However, apart from [Opris, 2025c; Opris, 2026a], we are not aware of any theoretical results on whether or how crossover can be beneficial in many-objective optimization. Moreover, these results are limited to Royal Road functions designed for specific crossover operators.

2 Preliminaries

Notation: For a finite set A, we write $|A|$ for its cardinality. For $n\in\mathbb{N}$, define $[n]:=\{1,\dots,n\}$ and 1^n the vector of length n with only ones, while 0^n is the corresponding vector with only zeros. Given a bit string $x\in\{0,1\}^n$, let $|x|_1$ and $|x|_0$ denote the number of ones and zeros in x, respectively. For $x,y\in\{0,1\}^n$ denote by $H(x,y)=\sum_{i=1}^n|x_i-y_i|$ the Hamming distance of x and y. We use ln to denote the natural logarithm. Let Y and Z be random variables taking values in $\mathbb{N}_0$. We say that Z *stochastically dominates* Y if $\Pr(Z\le c)\le\Pr(Y\le c)$ for all $c\ge 0$. Consider an m-objective function $f:\{0,1\}^n\to\mathbb{N}_0^m, x\mapsto(f_1(x),\dots,f_m(x))$, and let $f_{\max}:=\max\{f_j(x)\mid j\in[m],x\in\{0,1\}^m\}$ the maximum possible value of an objective. For two search points $x,y\in\{0,1\}^n$, we say that x *weakly dominates* y, denoted $x\succeq y$, if $f_i(x)\ge f_i(y)$ for all $i\in[m]$. If, in addition, at least one of these inequalities is strict, then x *dominates* y, written $x\succ y$. If neither $x\succeq y$ nor $y\succeq x$ holds, the two points are called *incomparable*. We say that a subset $S\subseteq\{0,1\}^n$ consists only of *mutually incomparable solutions* if every pair of elements in S is incomparable. A solution x is called *Pareto-optimal* if it is not dominated by any other search point in $\{0,1\}^n$, and we call the set $\{f(x)\mid x\in\{0,1\}^n$ is Pareto optimal$\}$ the *Pareto front*. For a population P_t and a fitness vector $v\in\mathbb{N}_0^m$, the *cover number* $c_t(v)$ is defined as the number of individuals $x\in P_t$ with $f(x)=v$. We say that v is *covered* if $c_t(v)\ge 1$.

The NSGA-III algorithm: The NSGA-III algorithm ([Deb and Jain, 2014]) with crossover probability $p_c\in[0,1)$ and

Algorithm 1: NSGA-III on an m-objective function f with population size μ and crossover probability p_c

```
Initialise P_0 ~ Unif(({0,1}^n)^μ)
for t := 0 to ∞ do
    Initialise Q_t := ∅
    for i = 1 to μ/2 do
        Select a_1, a_2 from P_t uniformly at random
        Select u ~ Unif([0,1])
        if u ≤ p_c then
            Create y_1 by uniform crossover on a_1, a_2
            Create y_2 by uniform crossover on a_1, a_2
        else
            Create y_1, y_2 as copies from a_1, a_2
        Create z_1, z_2 by standard bit mutation on y_1, y_2
        Update Q_t := Q_t ∪ {z_1, z_2}
    Set R_t := P_t ∪ Q_t
    Partition R_t into layers F_t^1, F_t^2, ..., F_t^k of
      non-dominated fitness vectors
    Find i* ≥ 1 such that Σ_{i=1}^{i*-1} |F_t^i| < μ and
      Σ_{i=1}^{i*} |F_t^i| ≥ μ
    Compute Y_t = ⋃_{i=1}^{i*-1} F_t^i
    Choose F̃_t^{i*} ⊂ F_t^{i*} such that |Y_t ∪ F̃_t^{i*}| = μ with
      Algorithm 2
    Create the next population P_{t+1} := Y_t ∪ F̃_{i*}^t
```

Algorithm 2: Selection procedure utilizing a set $\mathcal{R}_p$ of reference points for maximizing a function, including uniform selection.

```
Compute the normalisation f^n of f
Associate each x ∈ Y_t ∪ F_t^{i*} with its reference point
  rp(x) such that the distance between f^n(x) and the
  line through the origin and rp(x) is minimised
For each r ∈ R_p, set ρ_r := |{x ∈ Y_t | rp(x) = r}|
Initialise F̃_t^{i*} = ∅ and R' := R_p
while |F̃_t^{i*}| < μ/2 do
    Determine r_min ∈ R' such that ρ_{r_min} is minimal
      (where ties are broken randomly)
    Determine x_{r_min} ∈ F_t^{i*} \ F̃_t^{i*} which is associated
      with r_min and minimises the distance between
      the vectors f^n(x_{r_min}) and r_min (where ties are
      broken randomly)
    if x_{r_min} exists then
        F̃_t^{i*} = F̃_t^{i*} ∪ {x_{r_min}}
        ρ_{r_min} = ρ_{r_min} + 1
        if |Y_t| + |F̃_t^{i*}| = μ then
            return F̃_t^{i*}
    else
        R' = R' \ {r_min}
Select μ − |Y_t| − μ/2 distinct individuals uniformly at
  random from F_t^{i*} \ F̃_t^{i*}, and add them to F̃_t^{i*}
return F̃_t^{i*}
```

even population size μ is presented in Algorithm 1. Initially, a population P_0 of size μ is generated by selecting μ individuals uniformly at random from $\{0,1\}^n$. In each generation t, an offspring population Q_t of size μ is created by performing the following operations $\mu/2$ times. First, two parents a_1 and a_2 are selected uniformly at random from P_t. Then, *uniform crossover* is applied to (a_1, a_2) two times with probability p_c to produce two intermediate solutions y_1 and y_2. That is, for creating one solution, and for each position $i \in [n]$ independently, the entry from a_1 is taken with probability $1/2$, and otherwise the entry from a_2. If uniform crossover is not applied, y_1 and y_2 are exact copies of a_1 and a_2. Finally, two offspring z_1 and z_2 are generated by applying *standard bit mutation* to y_1 and y_2, that is, 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 and R_t is updated by partitioning R_t into layers $F_t^1, F_t^2, \dots$ using the *non-dominated sorting algorithm* [Deb *et al.*, 2002] where F_t^1 consists of all non-dominated individuals, and F_t^i for $i > 1$ of individuals only dominated by those from $F_t^1, \dots, F_t^{i-1}$. Then the critical rank i^* with $\sum_{i=1}^{i^*-1}|F_t^i| < \mu$ and $\sum_{i=1}^{i^*}|F_t^i| \geq \mu$ is determined (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^*). All individuals with a lower rank than i^* are included in P_{t+1}, while the remaining individuals are selected from $F_t^{i^*}$ using Algorithm 2. Hereby, a normalized objective function f^n is computed and then each individual with rank at most i^* is associated with reference points. For the first, we use the normalization procedure from [Wietheger and Doerr, 2023] which can be also used for maximization problems as shown in [Opris *et al.*, 2024a]. We omit detailed explanations as they are not needed for our purposes. For an m-objective function $f\colon \{0,1\}^n \to \mathbb{N}_0^m$, the normalized fitness vector $f^n(x) := (f_1^n(x), \dots, f_m^n(x))$ of a search point x is computed as

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

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

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

where $p \in \mathbb{N}$ is a parameter one can choose according to the fitness function f. These are uniformly distributed on the simplex determined by the unit vectors $(1,0,\ldots,0)^\intercal,(0,1,\ldots,0)^\intercal,\ldots,(0,0,\ldots,1)^\intercal$.
Then, if the number of all individuals already chosen for the critical layer is at most $\mu/2$, 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, or $|F_t^{i^*}| = \mu/2$, the selection ends. If after this selection procedure still $|Y_t| + |\tilde{F}_t^{i^*}| < \mu$, then the remaining $\mu - |Y_t| - |\tilde{F}_t^{i^*}| = \mu - |Y_t| - \mu/2 = \mu/2 - |Y_t|$ individuals from $F_{i^*} \setminus \tilde{F}_t^{i^*}$ are chosen uniformly at random. Such a uniform selection strategy has been shown to be successful, since it enables movement on fitness neutral environments (called *plateaus*), which helps to build up and preserve population diversity [Opris *et al.*, 2024b; Dang *et al.*, 2017; Doerr and Qu, 2023a]. The following result can be formulated and proven as in [Opris, 2025a]. For completeness, we provide a proof in the appendix.

Lemma 1. *Consider NSGA-III on an m-objective function f with Pareto front F, and assume that $\varepsilon_{\mathrm{nad}} \geq f_{\max}$. Let $\mathcal{R}_p$ denote a set of reference points, with $p \geq 2m^{3/2} f_{\max}$. Let P_t be the population at iteration t. Let S be a maximum set of mutually incomparable solutions, let $v \in F$, and $0 \leq \alpha \leq \lfloor \mu/(2|S|) \rfloor$. Then if $c_t(v) \geq \alpha$ then also $c_{t+1}(v) \geq \alpha$.*

The many-objective m-OJZJ$_k$ benchmark: This benchmark has been defined the first time in [Zheng and Doerr, 2024] and is defined as follows, where $m\text{-OJZJ}_k := (f_1,\ldots,f_m)$, m is even, and n is divisible by $m/2$ (see also [Doerr and Qu, 2022] for the bi-objective version). For $2 \leq k \leq 2n/m$ the $m\text{-OJZJ}_k = \big(f_1(x),\ldots,f_m(x)\big)$ is defined as

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

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

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

if $j \in [1,\ldots,m]$ is even. We often call k *gap size*. For every objective there are $2n/m+1$ different values and $f_{\max} = k + 2n/m$. The Pareto front F of m-OJZJ$_k$ is $\{(\ell_1, 2k+2n/m-\ell_1,\ldots,\ell_{m/2}, 2k+2n/m-\ell_{m/2}) \mid \ell_1,\ldots,\ell_{m/2} \in \{k,2k,2k{+}1,\ldots,2n/m{-}1,2n/m,2n/m{+}k\}\}$, has cardinality $(2n/m-2k+3)^{m/2}$ for $k \leq n/m$, and a maximum set S of mutually incomparable solutions satisfies $|F| \leq |S| \leq (2n/m+1)^{m/2}$ (see [Zheng and Doerr, 2024] for proofs).

For $v \in F$ we introduce the following notation: Denote by $r_v \in \{-1,0,1\}^{m/2}$ the $(m/2)$-dimensional vector with $(r_v)_j = -1$ if $v_{2j-1} = k$, $(r_v)_j = 1$ if $v_{2j-1} = 2n/m + k$, and $(r_v)_j = 0$ if $2k \leq v_{2j-1} \leq 2n/m$. All search points x with $f(x) \in \{v \in \mathbb{N}_0^m \mid r_v = s\}$ for $s \in \{-1,0,1\}^{m/2}$ satisfy $x^j = 0^{m/2}$ if $(r_v)_j = -1$, $x^j = 1^{m/2}$ if $(r_v)_j = 1$, and $k \leq |x^j|_1 \leq 2n/m - k$ if $(r_v)_j = 0$ for all $j \in [m/2]$.

3 An Upper Bound Without Crossover

To compare the performance of NSGA-III with and without crossover, we first generalize a result from [Opris, 2025a] for NSGA-III on m-OJZJ$_k$ to an arbitrary number of objectives. This generalization requires a much more refined proof, since for large m the number of Pareto-optimal points can be exponentially. Interestingly, the runtime bound from [Opris, 2025a] carries over directly to this setting in terms of generations. In addition, we provide an upper bound on the runtime that holds with high probability. Our results are formulated for arbitrary crossover probabilities $p_c \in [0,1)$, as parts of the analysis are later reused for the crossover case. However, the analysis here relies only on mutation steps.

Theorem 2. *Consider NSGA-III on $f := m$-OJZJ$_k$ for $2 \leq k \leq n/(2m)+1$, $(1+2n/m)^{m/2} \leq \mu/2 \leq n^{o(n)}$, crossover probability $p_c \in [0,1)$, and a number m of objectives with $2 \leq m \leq n/2$. Further, assume the same conditions as in Lemma 1, and let $\delta > 0$. Then with probability at least $1 - e^{-(\delta-3)m} - O(1/n^4)$ the number of generations until the whole Pareto front F is covered is at most*

$$O\left(\frac{m^2 n \ln(n/m)}{1-p_c} + \frac{\delta m n^k (1+2n/m)^{m/2}}{\mu(1-p_c)}\right).$$

The expected number of generations is

$$O\left(\frac{m^2 n \ln(n/m)}{1-p_c} + \frac{m n^k (1+2n/m)^{m/2}}{\mu(1-p_c)}\right).$$

Proof. At first we prove that, with probability at most $e^{-\Omega(n^2)}$, there exists no Pareto optimal individual x with $k \leq |x_j|_1 \leq 2n/m - k$ for all $j \in [m/2]$ after initialization. Then, with probability $1 - e^{-\Omega(n^2)}$, a single generation suffices to create such a Pareto optimal individual since any individual x can be created with probability at least n^{-n}, independently of whether crossover is executed or not. This implies that the expected number of generations for creating a Pareto optimal individual is at most $1 + n^n \cdot e^{-\Omega(n^2)} = 1 + o(1)$.

Lemma 3. *With probability at most $e^{-\Omega(n^2)}$, there exists no Pareto optimal individual x with $k \leq |x_j|_1 \leq 2n/m - k$ for all $j \in [m/2]$ after initialization.*

After a successful initialization, we explore search points x satisfying $r_{f(x)} = s$ for all possible $s \in \{-1,0,1\}^{m/2}$ in parallel. Specifically, we determine a suitable number of generations such that, for any fixed s, a search point x with $r_{f(x)} = s$ is found with high probability. We then apply a union bound over all $3^{m/2}$ possible s, which shows that, still with high probability, a corresponding x with $r_{f(x)} = s$ is found for every s within that time.

We fix an $s \in \{-1,0,1\}^{m/2}$ and define *Step j* as the process that starts when there is a point $y \in P_t$ with $(r_{f(y)})_i = s_i$ for all $i < j$ and $(r_{f(y)})_i = 0$ for $i \geq j$, and ends when a search point z is generated with $(r_{f(z)})_i = (r_{f(y)})_i$ for all $i \leq j$, and $(r_{f(z)})_i = 0$ for $i > j$. Note that Step 1 directly builds on a successful initialization (since then there exists a search point $z \in P_t$ with $r_{f(z)} = 0^{m/2}$), and that the desired z is created after Step $m/2$. For Step j we fix such a $y \in P_t$. Then in all future iterations there is always a $y' \in P_t$ with $f(y') = f(y)$ by Lemma 1, and hence, we can never fall back in steps. Further, we define $B_j := \{v \in F \mid v_i = f_i(y) \text{ for every } i \in [m] \setminus \{2j-1, 2j\} \text{ and } v_{2j-1} \in \{2k, \dots, n-2k\}\}$. Note that $f(y) \in B_j$. Now we estimate the number of generations to finish Step j.

- At first we estimate the number X_j of generations until the complete set B_j is covered. Particularly, that for every $v \in B_j$, there is $x^* \in P_t$ with $f(x^*) = v$.
- After X_j generations, we estimate the number of generations Y_j until every Pareto optimal fitness vector $v \in B_j$ has a cover number of at least $\lfloor \mu/(2(1+2n/m)^{m/2}) \rfloor$.
- Finally, after $X_j + Y_j$ generations, estimate the number of generations Z_j until a desired z is created by possibly crossing the fitness valley of size k if necessary.

Then the total number of generations until a desired x is created is stochastically dominated by $\sum_{j=1}^{m/2}(X_j + Y_j + Z_j)$. We now derive tail bounds for the random variables $X := \sum_{j=1}^{m/2} X_j$, $Y := \sum_{j=1}^{m/2} Y_j$, and $Z := \sum_{j=1}^{m/2} Z_j$ separately in three consecutive lemmas. We only provide some proof sketches due to space restrictions and similar ideas to the considerations from [Opris, 2025a]. Their full proofs can be found in the appendix.

Lemma 4. *Fix $\beta > 0$. Then for $j \in [m/2]$ and n sufficiently large,*

$$\Pr\left(X_j \geq \frac{4+16m}{1-p_c} \cdot n \cdot \ln\Big(\frac{2n}{m}\Big)\right) \leq \left(\frac{2n}{m}\right)^{-4m}.$$

Particulary,

$$\Pr\left(X \geq \frac{4+16m}{1-p_c} \cdot \frac{mn}{2} \cdot \ln\Big(\frac{2n}{m}\Big)\right) \leq \frac{m}{2} \cdot \left(\frac{2n}{m}\right)^{-4m}.$$

We fix an uncovered Pareto-front fitness vector v and bound the number of generations until it is covered with high probability. A union bound over all $v \in B_j$ then yields a high-probability bound for covering all such vectors, which extends to X via another union bound by multiplying with $m/2$. We next estimate Y.

Lemma 5. *For $j \in [m/2]$ we have*

$$\Pr\left(Y_j \geq \frac{3207n}{1-p_c}\right) \leq 2n^2 e^{-4n}.$$

Particulary,

$$\Pr\left(Y \geq \frac{3207nm}{2(1-p_c)}\right) \leq mn^2 e^{-4n}.$$

The idea behind proving that each fitness vector from B_j attains the desired cover number after $\Omega(n/(1-p_c))$ generations is again based on a parallelization argument. We split the proof into two phases. In the first phase, we determine an expected number of generations such that the cover number of a fixed $v \in B_j$ is at least $\Omega(n)$ with high probability, which follows from repeatedly cloning an individual x with $f(x) = v$. In the second phase, we show that the desired cover number $c_t(v)$ is reached by applying a classical Chernoff bound to the number of newly created individuals in a single generation, which is $\Omega(n)$ in expectation. By a union bound, we obtain the claim for all $v \in B_j$, and finally also for Y. Next, we extimate Z.

Lemma 6. *For $\gamma := \lfloor \mu/(2(1+2n/m)^{m/2}) \rfloor$ and $j \in [m/2]$ the random variable Z_j is stochastically dominated by a geometrically distributed random variable Z'_j with success probability $\sigma_k := \frac{(1-p_c)\gamma/(2en^k)}{1+(1-p_c)\gamma/(2en^k)}$. Additionally, for $\delta > 1$*

$$\Pr\left(Z \geq \frac{m(1+8\delta)}{2}\Big(1 + \frac{2en^k}{(1-p_c)\gamma}\Big)\right) \leq e^{-\delta m}.$$

We estimate Z_j pessimistically as the number of generations required to select an individual with $2n/m - k$ ones (or, symmetrically, $2n/m - k$ zeros) in block j, of which there are at least $\lfloor \mu/(2(1+2n/m)^{m/2}) \rfloor$ many, and then flip k specific bits in block j to cross the fitness valley. Again, we apply a union bound to estimate Z.

Finally, we are in a position to apply a union bound to the random variables X, Y and Z. Using Lemmas 4, 5, and 6, respectively, we obtain for every $\delta \geq 1$ and

$$\begin{aligned} K(k,m,n,p_c,\delta) &:= \frac{4+16m}{1-p_c} \cdot \frac{mn}{2} \cdot \ln\Big(\frac{2n}{m}\Big) + \frac{3207nm}{2(1-p_c)} \\ &\quad + \frac{m(1+8\delta)}{2}\Big(1 + \frac{2en^k}{(1-p_c)\gamma}\Big) \\ &= O\left(\frac{m^2 n \ln(n/m)}{1-p_c} + \frac{\delta m n^k (1+2n/m)^{m/2}}{(1-p_c)\mu}\right) \end{aligned}$$

the inequality

$$\begin{aligned} &\Pr(X + Y + Z \geq K(k,m,n,p_c,\delta)) \\ &\leq m/2 \cdot (2n/m)^{-4m} + mn^2 e^{-4n} + e^{-\delta m} =: p(m,n,\delta). \end{aligned}$$

So after $K(k,m,n,p_c,\delta)$ generations, a desired x with $r_{f(x)} = s$ is created with probability at most $p(m,n,\delta)$, after a successful initialization. A union bound on at most $3^{m/2}$ such possible x with different $r_{f(x)}$-values shows that with probability at most (due to $2n/m \geq 4$)

$$3^{\frac{m}{2}} p(m,n,\delta) \leq m(2n/m)^{-3m}/2 + mn^2 e^{-3n} + e^{-(\delta-3)m}$$

that for each $s \in \{-1,0,1\}^{m/2}$ there is x with $r_{f(x)} = s$. Suppose that this happens. Then we cover the Pareto front. Fix an uncovered $v \in F$. In an analogous way to the argument above, the number of generations required to cover v is stochastically dominated by X. This is because there already exists a search point $y \in P_t$ with $r_{f(y)} = r_v$ and hence, for every block $j \in [m/2]$ with $(r_v)_j = 0$, the time to create a search point z satisfying $(f_{2j-1}(z), f_{2j}(z)) = (v_{2j-1}, v_{2j})$, while not changing all remaining blocks, is stochastically dominated by X_j. By Lemma 4, we obtain

$$\Pr\left(X \geq \frac{(4+16m)mn\ln(2n/m)}{2(1-p_c)}\right) \leq \left(\frac{2n}{m}\right)^{-4m}.$$

By a union bound over all such possible search points, we obtain that the whole Pareto front is covered in an additional amount of $(4+16m)\cdot mn/2\cdot\ln(2n/m)/(1-p_c)$ generations with probability at most $(1+2n/m)^{m/2}\cdot(2n/m)^{-4m} \leq (2n/m)^{-2m}$. Hence, for $p := p(m,n,\delta)$, the entire Pareto front is covered within

$$K(k,m,n,p_c,\delta) + \frac{4+16m}{1-p_c}\cdot\frac{mn}{2}\cdot\ln\Big(\frac{2n}{m}\Big) = O\left(\frac{m^2 n\ln(n/m)}{1-p_c} + \frac{\delta m n^k(1+2n/m)^{m/2}}{(1-p_c)\mu}\right)$$

generations with probability at least

$$1 - 3^{m/2}p - (2n/m)^{-2m} = 1 - e^{-(\delta-3)m} - O(1/n^4)$$

where we also used that the functions $h_1 :]0, n/2[\to \mathbb{R}, x \mapsto (2n/x)^{-2x}$, and $h_2 :]0, n/2[\to \mathbb{R}, x \mapsto x/2\cdot(2n/x)^{-2x}$, are strictly monotone decreasing (see Lemma **??** in the appendix). The latter propability also includes the event that in the first generation a successful initialization happens.
Now it remains to estimate the expected time until the whole Pareto front is covered. Here we assume $\delta = 4$. If, after a successful initialization, the Pareto front is not covered after $O(m^2 n\ln(n/m)/(1-p_c) + mn^k(1+2n/m)^{m/2}/((1-p_c)\mu))$ generations (which happens with probability at most $e^{-m} + O(1/n^4)$), then we repeat all the arguments from above. The expected number of such periods is $1/(1-e^{-m} - O(1/n^4)) = O(1)$, concluding the proof of Theorem 2. □

4 An Upper Bound with Crossover

Now we show that NSGA-III with crossover can be much more efficient when optimizing m-OJZJ$_k$ than without, especially for large k. The reason is that a pair x, y with $|x^j|_1 = |y^j|_1 = 2n/m - k$ and maximum Hamming distance $H(x^j, y^j) = 2k$ can be created via mutation within k generations, and then a recombination of those x^j and y^j leads to the all one string in block j with probability $\Omega(1/4^k)$. This probability is by a factor $n^k/4^k$ larger than the probability to cross the fitness valley in block j via mutation.

Theorem 7. *Consider NSGA-III on $f := m$-OJZJ$_k$ for $2 \leq k \leq n/(2m)+1$, $(1+2n/m)^{m/2} \leq \mu/2 \leq n^{o(n)}$, crossover probability $p_c \in (0,1)$, and a number m of objectives with $2 \leq m \leq n/2$. Further, assume the same conditions as in Lemma 1, and let $\delta > 0$. Then, with probability at least $1 - e^{-(\delta-3)m} - O(1/n^4)$, the number of generations until the whole Pareto front F is covered is at most*

$$O(k\delta m^{k+1} n^k (80e)^k \mu/(p_c k!(1-p_c)^k)).$$

The expected number of generations is

$$O(km^{k+1} n^k (80e)^k \mu/(p_c k!(1-p_c)^k)).$$

Proof. We closely follow the ideas used in the proof of Theorem 2. For the initialization, with probability $1 - e^{-\Omega(n^2)}$, a single generation suffices to initialize an individual x with $k \leq |x^j|_1 \leq 2n/m - k$ for all $j \in [m/2]$. The expected number of generations is also $1 + o(1)$. Now we define Step j and the corresponding set B_j as in the proof of Theorem 1. Let X_j denote the number of generations until, for every $v \in B_j$, there exists a search point $x^* \in P_t$ with $f(x^*) = v$. After these X_j generations, let Z_j^{cross} be the number of additional generations until a desired z is created. As in Lemma 4, we see that

$$\Pr\left(X \geq \frac{4+16m}{1-p_c}\cdot\frac{mn}{2}\cdot\ln\Big(\frac{2n}{m}\Big)\right) \leq \frac{m}{2}\left(\frac{2n}{m}\right)^{-4m}.$$

Now we estimate Z. To estimate Z_j, we consider $k+1$ consecutive generations. In the first k generations, we create two suitable individuals y_1, x_2 with maximum Hamming distance in block j which can then be used for recombination in the $(k+1)$-th generation.

Lemma 8. *For $j \in [m/2]$ the random variable Z_j is stochastically dominated by $(k+1)Z_j^{cross}$ where Z_j^{cross} is a geometrically distributed random variable with success probability $\sigma_k := \frac{1}{10^k}\frac{p_c/(2\cdot 4^{k+1}\mu)}{1+p_c/(2\cdot 4^{k+1}\mu)}\prod_{\ell=1}^{k}\frac{(1-p_c)\ell/(2emn)}{1+(1-p_c)\cdot\ell/(2emn)}$. Additionally, for $\delta > 1$,*

$$\Pr(Z \geq (k+1)(1+8\delta)m/(2\sigma_k)) \leq e^{-\delta m}.$$

Proof. As in the proof of Lemma 4, for the desired search point z to create, we may assume that $z_j = 1^{m/2}$. Let $s := r_{f(z)}$. Consider a sequence of $k+1$ generations as follows: Within the first k generations, one may create a suitable pair of individuals x_1, x_2 such that $|x_1^j|_1 = |x_2^j|_1 = 2n/m - k$ and the Hamming distance in block j is maximized, particularly $H(x_1^j, x_2^j) = 2k$, while the individuals coincide in all other blocks. Moreover, for all $i < j$, we have $(r_{f(x_1)})_i = (r_{f(x_2)})_i = s_i$ and we have $(r_{f(x_1)})_i = (r_{f(x_2)})_i = 0$ for $i > j$. In the $(k+1)$-th generation, perform uniform crossover on x_1 and x_2, and do not flip any bit during mutation, to create a suitable search point z with $z^j = 1^{2n/m}$, particularly $(r_{f(z)})_i = s_i$ for all $i \leq j$, and $(r_{f(z)})_i = 0$ for all $i > j$. So for generation $\ell \leq k$, assume that there are two individuals $x_{\ell,1}, x_{\ell,2}$ with $|x_{\ell,1}^j|_1 = |x_{\ell,2}^j|_1 = 2n/m - k$, $H(x_{\ell,1}^j, x_{\ell,2}^j) = 2(\ell-1)$, where for $\ell = 1$, the individuals $x_{\ell,1}, x_{\ell,2}$ are the same. They also coincide in all the other blocks, where $(r_{f(x_{\ell,1})})_i = s_i$ for $i < j$, and $(r_{f(x_{\ell,1})})_i = 0$ for $i > j$ is satisfied. Then in one trial, one may choose x_1 as a parent (prob. at least $1/\mu$), omit the crossover, and flip exactly two bits, namely, one from 1 to 0 and one from 0 to 1 in block j, both of which do not contribute to $H(x_1, x_2)$. Note that there are $k - (\ell - 1)$ such zeros, and $2n/m - 2(\ell-1)$ such ones. Hence, this happens with probability at least $(1-p_c)(k-\ell+1)/n\cdot(2n/m-(\ell-1)-k)/n\cdot(1-1/n)^{n-2} \geq (1-p_c)(k-\ell+1)/n\cdot(2n/m-2k+1)/(en) \geq (1-p_c)(k-\ell+1)/n\cdot(n/m-1)/(en) \geq (1-p_c)(k-\ell+1)/(2emn) =: q_\ell$, and in one generation with probability at least

$$1 - \left(1 - \frac{q_\ell}{\mu}\right)^{\mu/2} \geq \frac{q_\ell/2}{1+q_\ell/2}.$$

These x_1 and x_2 then both survive with probability at least $1/10$: If $|F_t^1| \leq \mu$ where $|F_t^1|$ is the number of individuals with rank one, then both survive with probability 1. Otherwise, $\mu < |F_t^1| \leq 2\mu$, and they survive with probability at least (when both are not selected within the first $\mu/2$ iterations of the reference point scheme in Algorithm 2) $\binom{\mu/2}{2}/\binom{|F_t^1|-\mu/2}{2} \geq \binom{\mu/2}{2}/\binom{3\mu/2}{2} \geq 1/10$ for n sufficiently large. Hence, in a sequence of k generations, the probability that there is such a pair x_1, x_2 of individuals is at least $1/10^k \prod_{\ell=1}^k (q_\ell/2)/(1+q_\ell/2)$. For the subsequent $(k+1)$-th generation, there are two such individuals x_1, x_2. Now, choose x_1, x_2 as parents (prob. at least $1/\mu^2$), perform uniform crossover to create y' with $(r_{f(y')})_i = s_i$ for all $i \leq j$, particularly $(y')_i = 1^{2n/m}$ and omit mutation (prob. $p_c/4^k \cdot (1-1/n)^n \geq p_c/4^{k+1}$). Hence, such a desired y' is generated in this $(k+1)$-th generation with probability at least

$$1 - \left(1 - \frac{p_c}{4^{k+1}\mu^2}\right)^{\mu/2} \geq \frac{p_c/(2 \cdot 4^{k+1}\mu)}{1 + p_c/(2 \cdot 4^{k+1}\mu)} =: p^{\text{cross}},$$

which survives by Lemma 1. Hence, the probability is at least $\sigma_k = \frac{p^{\text{cross}}}{10^k} \cdot \prod_{\ell=1}^k \frac{q_\ell/2}{1+q_\ell/2}$ to create the desired y' within $k+1$ generations. Hence, the random variable $Z = \sum_{j=1}^{m/2} Z_j$, is indeed stochastically dominated by $(k+1)Z^*$, where $Z^* = \sum_{j=1}^{m/2} Z_j^*$ is an independent sum of geometrically distributed random variables Z_j^* with success probability σ_k. Its expectation is $m/(2\sigma_k)$. Hence, we have that

$$\begin{aligned}\Pr(Z \geq (k+1)\mathrm{E}[Z^*] + (k+1)\lambda) &= \Pr(Z^* \geq \mathrm{E}[Z^*] + \lambda) \\ &\leq \exp\left(-\frac{1}{4}\min\left\{\frac{\lambda^2}{d}, \lambda\sigma_k\right\}\right),\end{aligned}$$

for all $\lambda \geq 0$ where $d := m/(2\sigma_k^2)$. For $\lambda = 4m\delta/\sigma_k$, where $\delta \geq 1$ is arbitrary, we observe that $\Pr(Z \geq (k+1) \cdot (1 + 8\delta)m/(2\sigma_k)) \leq e^{-\delta m}$ proving the lemma. □

Now we take a union bound to obtain with Lemmas 4 and 8 for $W := X + Z$

$$\begin{aligned}\Pr\left(W \geq \frac{(4+16m)mn\ln(2n/m)}{2(1-p_c)} + \frac{(k+1)(1+8\delta)m}{2\sigma_k}\right) \\ \leq (2n/m)^{-4m} + e^{-\delta m}.\end{aligned}$$

Hence, after $O(nm\ln(n/m)/(1-p_c) + k\delta m 40^k \mu/p_c \cdot (2emn)^k/(k!(1-p_c)^k)) = O(k\delta m^{k+1} n^k (80e)^k \mu/(p_c k!(1-p_c)^k))$ generations, we see that a desired x with $r_{f(x)} = s$ is created with probability at most $(2n/m)^{-4m} + e^{-\delta m}$. A union bound over at most $3^{m/2}$ choices of x with different $r_{f(x)}$ values shows that, with probability at most $e^{-(3-\delta)m} + O(1/n^4)$, after this number of generations there exists for each $s \in -1, 0, 1$ an x with $r_{f(x)} = s$. Conditioned on this event, the final covering step proceeds as in Theorem 2 and succeeds with probability at least $1 - O(1/n^4)$ within an additional $O(m^2 n \ln(n/m)/(1-p_c))$ generations. Hence, with probability at least $1 - e^{-(\delta-3)m} - O(1/n^4)$, the whole Pareto front is covered in $O(k\delta m^{k+1} n^k (80e)^k \mu/(p_c k!(1-p_c)^k))$ generations. If this does not happen, we repeat the above arguments to obtain an expected number of $O(km^{k+1} n^k (80e)^k \mu/(p_c k!(1-p_c)^k))$ generations, where we use $\delta = 4$, proving the theorem. □

5 A Lower Bound for 4-OJZJ

To complement the upper bound for the crossover case, we also establish a lower bound for mutation only in the case $m = 4$ and for population sizes asymptotically smaller than n^k by a factor of n^2.

Theorem 9. *Consider NSGA-III on $f := 4\text{-OJZJ}_k$ for $2 \leq k \leq n/8 + 1$, $\mu \in O(n^{k-2})$, where crossover is turned off. Then the expected number of generations until the whole Pareto front is covered is at least $\Omega(n^k/\mu)$.*

Proof. By a classical Chernoff bound, with probability $1 - e^{-\Omega(n)} = 1 - o(1)$, each individual x satisfies $|x^j|_1 \in \{k, \ldots, n/2 - k\}$ for both $j \in \{1, 2\}$ after initialization. Suppose that this happens. Then in all future populations $P_1, P_2, \ldots$ there are only individuals x with $|x^j|_1 \in \{0, k, \ldots, n/2 - k, n/2\}$ for at least one $j \in \{1, 2\}$, since other individuals are dominated by every y with $|y^j|_1 \in \{k \ldots, n/2 - k\}$ for both $j \in \{1, 2\}$.

Lemma 10. *With probability at least $1/81 - o(1)$, we have that $C_t := \{s \in \{-1, 0, 1\}^2 \mid$ there is $x \in P_t$ with $r_{f(x)} = s\} \neq \{-1, 0, 1\}^2$ after $cn\ln(n)$ generations for a constant $c > 0$, and C_t can only increase by flipping either k specific ones or zeros in a block $j \in \{1, 2\}$.*

Assume the event from Lemma 10 occurs. Then flipping k specified ones or zeros within a block happens with probability at most $2n^{-k}$ per trial, so the expected number of trials is at least $(1 - o(1))n^k/2 = \Omega(n^k)$. This yields an expected number of $\Omega(n^k/\mu)$ generations. □

6 Conclusion

In this paper, we derived upper runtime bounds for the widely used NSGA-III algorithm on m-OJZJ$_k$ with an arbitrary number m of objectives, both with and without crossover, in expectation and with high probability. Notably, crossover can significantly speed up the runtime and even lead to exponential improvements in suitable parameter regimes, for example when the gap parameter k is large while the number of objectives and the population size are small. To complement these findings, we also established a lower runtime bound for the case $m = 4$. We hope these results improve the understanding of crossover in MOEAs, especially for escaping local optima and improving search efficiency. Future work could investigate lower bounds for the crossover case, and for $m > 4$ if crossover is turned off. The latter is more challenging due to population dynamics, and we still lack a clear understanding of how non-Pareto-optimal individuals behave in the population. For crossover, the challenge is that knowledge about objective vectors alone are insufficient. One has to know also about the distribution of the genotypes of individuals. Another direction is to explore crossover in more complex many-objective settings, such as combinatorial optimization problems like shortest paths, or permutation spaces, like the many-objective flow-shop scheduling problem.

References

[Bi-wu Zhu *et al.*, 2025] Bi-wu Zhu, Zi wen Feng, Hao Jiang, Xiao Liu, Jian zhao Wu, Wen hui Liu, Fan Ye, Yu xin Lin, Peng cheng Guo, Cong chang Xu, and Luo xing Li. A strategy combining interpretable machine learning, nsga-iii optimization model and strengthening and toughing mechanism to predict microstructure for rolled az31 magnesium alloy sheets. *Journal of Materials Research and Technology*, 39:1028–1037, 2025.

[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.

[Campos Ciro *et al.*, 2016] Guillermo Campos Ciro, Frédéric Dugardin, Farouk Yalaoui, and Russell Kelly. A nsga-ii and nsga-iii comparison for solving an open shop scheduling problem with resource constraints. *IFAC-PapersOnLine*, 49(12):1272–1277, 2016.

[Dang *et al.*, 2017] Duc-Cuong Dang, Tobias Friedrich, Timo Kötzing, Martin S. Krejca, Per Kristian Lehre, Pietro S. Oliveto, Dirk Sudholt, and Andrew M. Sutton. Escaping local optima using crossover with emergent diversity. *IEEE Transactions on Evolutionary Computation*, 22:484–497, 2017.

[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 *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI 2023*, pages 12390–12398. AAAI Press, 2023.

[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.

[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 *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2023, pages 12399–12407. AAAI Press, 2023.

[Doerr *et al.*, 2012] Benjamin Doerr, Edda Happ, and Christian Klein. Crossover can provably be useful in evolutionary computation. *Theoretical Computer Science*, 425:17–33, 2012.

[Doerr *et al.*, 2013] Benjamin Doerr, Daniel Johannsen, Timo Kötzing, Frank Neumann, and Madeleine Theile. More effective crossover operators for the all-pairs shortest path problem. *Theoretical Computer Science*, 471:12–26, 2013.

[Doerr *et al.*, 2022] Benjamin Doerr, Omar El Hadri, and Adrien Pinard. The $(1 + (\lambda, \lambda))$ global SEMO algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'22)*, pages 520–528. ACM Press, 2022.

[Doerr *et al.*, 2025] Benjamin Doerr, Dimitri Korkotashvili, and Martin S. Krejca. Difficulties of the nsga-ii with the many-objective leadingones problem. *IEEE Transactions on Evolutionary Computation*, pages 1–1, 2025.

[Doerr *et al.*, 2026] Benjamin Doerr, Martin S. Krejca, and Milan Stankovic. Improved runtime guarantees for the SPEA2 multi-objective optimizer. In *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI 2026*, pages 36855–36863. AAAI Press, 2026.

[Gaeuman and Sutton, 2025] Liam Gaeuman and Andrew M. Sutton. A fixed-parameter tractable ga for data clustering. In *Proceedings of the 18th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*, FOGA '25, page 49–60. ACM, 2025.

[Ishibuchi *et al.*, 2014] Hisao Ishibuchi, Yuki Tanigaki, Hiroyuki Masuda, and Yusuke Nojima. Distance-based analysis of crossover operators for many-objective knapsack problems. In *Proseecings of the Parallel Problem Solving from Nature*, PPSN XIII, pages 600–610, 2014.

[Jansen and Wegener, 2002] Thomas Jansen and Ingo Wegener. On the analysis of evolutionary algorithms—a proof that crossover really can help. *Algorithmica*, 34(1):47–66, 2002.

[Jansen and Wegener, 2005] Thomas Jansen and Ingo Wegener. Real royal road functions—where crossover provably is essential. *Discrete Applied Mathematics*, 149:111–125, 2005.

[Kötzing *et al.*, 2011] Timo Kötzing, Dirk Sudholt, and Madeleine Theile. How crossover helps in pseudo-boolean optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'11)*, pages 989–996. ACM Press, 2011.

[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.

[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.

[Opris and Antipov, 2026] Andre Opris and Denis Antipov. Parent selection mechanisms in elitist crossover-based algorithms, 2026. to appear at GECCO 2026.

[Opris *et al.*, 2024a] 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 *et al.*, 2024b] Andre Opris, Johannes Lengler, and Dirk Sudholt. A tight $O(4k/p_c)$ runtime bound for a $(\mu + 1)$ ga on JUMP_k for realistic crossover probabilities. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'24)*, page 1605–1613. ACM Press, 2024.

[Opris *et al.*, 2025] Andre Opris, Sebastian Sonntag, and Dirk Sudholt. A royal road function for permutation spaces: an example where order crossover is provably essential. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '25, page 1631–1640. ACM, 2025.

[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 *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI 2025, pages 8903–8911. ijcai.org, 2025.

[Opris, 2025b] Andre Opris. A first runtime analysis of the paes-25: An enhanced variant of the pareto archived evolution strategy. In *Proceedings of the 18th ACM/SIGEVO Conference on Foundations of Genetic Algorithms*, FOGA '25, page 202–213, New York, NY, USA, 2025. ACM.

[Opris, 2025c] 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.

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

[Opris, 2026b] Andre Opris. Towards a rigorous understanding of the population dynamics of the NSGA-III: Tight runtime bounds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, AAAI 2026, pages 37125–37133. AAAI Press, 2026.

[Pavai and Geetha, 2016] G Pavai and V Geetha. A survey on crossover operators. *ACM Comput. Surv.*, 49(4), December 2016.

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

[Qian *et al.*, 2020] Chao Qian, Chao Bian, and Chao Feng. Subset selection by pareto optimization with recombination. In *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 2408–2415. AAAI Press, 2020.

[Rashmi *et al.*, 2025] Sharma Rashmi, Jesubalan Naveen G., and Anurag S. Rathore. Integrating ensemble nsga-ii for multi-objective process optimization: Refolding of proinsulin as a case study. *Biotechnology and Bioengineering*, 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 *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN '24)*, LNCS, page to appear. Springer, 2024.

[Sharma *et al.*, 2021] Shreya Sharma, Julian Blank, Kalyanmoy Deb, and Bijaya Ketan Panigrahi. Ensembled crossover based evolutionary algorithm for single and multi-objective optimization. In *2021 IEEE Congress on Evolutionary Computation (CEC)*, pages 1439–1446, 2021.

[Sutton and Lee, 2024] Andrew Sutton and Jiwon Lee. Evolving populations of solved subgraphs with crossover and constraint repair. In *Proceedings of the International Conference on Parallel Problem Solving from Nature (PPSN '24)*, page 133–148. Springer, 2024.

[Vijai and P., 2025] Praveen Vijai and Bagavathi Sivakumar P. A hybrid multi-objective optimization approach with NSGA-II for feature selection. *Decision Analytics Journal*, 14:100550, 2025.

[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.org, 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.

[Zheng and Doerr, 2024] 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.