

Runtime Analysis for the NSGA-II: Provable Speed-Ups From Crossover

Benjamin Doerr, Zhongdi Qu

Laboratoire d’Informatique (LIX), Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, Palaiseau, France
lastname@lix.polytechnique.fr, d.q.1124@gmail.com

Abstract

Very recently, the first mathematical runtime analyses for the NSGA-II, the most common multi-objective evolutionary algorithm, have been conducted. Continuing this research direction, we prove that the NSGA-II optimizes the OneJumpZeroJump benchmark asymptotically faster when crossover is employed. Together with a parallel independent work by Dang, Opris, Salehi, and Sudholt, this is the first time such an advantage of crossover is proven for the NSGA-II. Our arguments can be transferred to single-objective optimization. They then prove that crossover can speed up the $(\mu + 1)$ genetic algorithm in a different way and more pronounced than known before. Our experiments confirm the added value of crossover and show that the observed advantages are even larger than what our proofs can guarantee.

Introduction

The theory of randomized search heuristics (Auger and Doerr 2011) has greatly improved our understanding of heuristic search, in particular, via mathematical runtime analyses. Due to the complicated nature of the stochastic processes describing runs of these algorithms, mostly very simple, often synthetic, heuristics could be analyzed so far. Very recently, however, Zheng, Liu, and Doerr (2022) conducted a runtime analysis for the NSGA-II (Deb et al. 2002), the most common evolutionary multi-objective (EMO) algorithm (47000 citations on Google scholar). This work was quickly followed up in different directions (Zheng and Doerr 2022a; Bian and Qian 2022; Doerr and Qu 2022b; Zheng and Doerr 2022b; Doerr and Wietheger 2022). The majority of these studies regards a simplified version of the NSGA-II that does not use crossover. Only in (Bian and Qian 2022) an NSGA-II with crossover is analyzed, but the runtime guarantees shown there were not better than those shown before for the mutation-only NSGA-II.

In this work, we conduct the (together with the parallel work (Dang et al. 2023)) first mathematical runtime analysis of the NSGA-II with crossover that proves runtime guarantees asymptotically stronger than those known for the NSGA-II without crossover. To this aim, we regard the ONEJUMPZEROJUMP benchmark (Doerr and Zheng 2021),

which is a bi-objective version of the classic JUMP benchmark intensively studied in the analysis of single-objective search heuristics. The mutation-based NSGA-II with population size N at least four times the size of the Pareto front computes the Pareto front of this problem in expected time (number of function evaluations) $O(Nn^k)$, where n is the length of the bit-string representation and k is the jump size, a difficulty parameter of the problem (Doerr and Qu 2022b). The authors say that they believe this bound to be tight, but prove a lower bound of $\Omega(n^k)$ only. That work includes preliminary experiments that show performance gains from crossover, but without proving them or giving additional insights on why they arise.

Our main result is that the NSGA-II using crossover with a constant rate solves this problem in expected time $O(N^2n^k/\Theta(k)^k)$. This is faster than the previous bound and the $\Omega(n^k)$ lower bound when $k \geq c \log(N)/\log \log(N)$, where c is a suitable constant. The key to proving this performance advantage is a careful analysis of how the NSGA-II arrives at having substantially different solutions with the same objective value in its population. This appears to be an advantage of the NSGA-II over algorithms previously analyzed mathematically such as the SEMO or global SEMO, where for each objective value at most one solution is kept in the population.

As a side result, we observe that our arguments for proving diversity in the population can also be used in single-objective optimization. This way we show that the $(\mu + 1)$ genetic algorithm (GA) – without additional diversity mechanisms or other adjustments favoring crossover – optimizes the jump function with encoding length n and jump size k in time $O(\mu^{k+1}n^k/\Theta(k)^k)$. This result holds for all n , $2 \leq k = o(\sqrt{n})$, and $2 \leq \mu$. For many parameter values, this compares favorably with the known $\Omega(n^k)$ runtime of the mutation-only analog of the $(\mu + 1)$ GA, the so-called $(\mu + 1)$ EA. Different from the previous analyses of the $(\mu + 1)$ GA on JUMP in (Dang et al. 2018; Doerr et al. 2022), our result shows super-constant speed-ups already for constant population sizes and our speed-ups increase with the problem difficulty k .

Our experimental results confirm the advantages of crossover for both the NSGA-II and the $(\mu + 1)$ GA. The advantages observed are even stronger than what our proofs guarantee. We note that it is not uncommon that mathemat-

ical runtime analyses cannot prove all advantages of an algorithm visible in experiments. In return, they give proven results and explanations on how the advantages arise.

Overall, our results give a strong mathematical justification for using the NSGA-II with crossover as proposed in the original work (Deb et al. 2002). Our new analysis methods promise to be applicable also to other randomized search heuristics using crossover, as demonstrated briefly on the $(\mu + 1)$ GA.

We note that a similar result appeared in the parallel independent work (Dang et al. 2023). There a problem was constructed for which the mutation-based NSGA-II has an exponential runtime, but with crossover the runtime becomes polynomial. From the main proof arguments, it appears unlikely that this result has a single-objective analog.

Previous Works

Since this work conducts a mathematical runtime analysis of a classic, crossover-based EMO algorithm, let us briefly describe the state of the art in runtime analysis with respect to EMO algorithms and crossover.

Mathematical runtime analyses have a long history in the field of heuristic search, cf. (Sasaki and Hajek 1988; Bäck 1993; Gutjahr 2008; Neumann and Witt 2015; Lissovoi and Oliveto 2018; Lissovoi, Oliveto, and Warwicker 2019) or the textbooks (Neumann and Witt 2010; Auger and Doerr 2011; Jansen 2013; Doerr and Neumann 2020). While often restricted to simple algorithms and problems, the mathematical nature of these works has led to many deep insights that could not have been obtained with empirical methods.

The mathematical runtime analysis of EMO algorithms was started in (Laumanns et al. 2002; Giel 2003; Thierens 2003). As in the theory of single-objective heuristics, these and most subsequent works analyzed simple synthetic algorithms. At the last AAI conference, the first runtime analysis of the NSGA-II algorithm (Deb et al. 2002), the most common EMO algorithm in practice, was presented (Zheng, Liu, and Doerr 2022). It proved that the NSGA-II with suitable population size can find the Pareto front of the ONEMINMAX and LOTZ benchmarks in the same asymptotic runtimes that were shown previously for the SEMO and GSEMO synthetic algorithms. This result is remarkable in that the NSGA-II has much more complex population dynamics than the (G)SEMO. In particular, it can lose desirable solutions due to the complex selection mechanism building on non-dominated sorting and crowding distance (and this was proven to happen consistently when the population size is equal to the size of the Pareto front (Zheng, Liu, and Doerr 2022)).

Zheng and Doerr (2022a) proved that the NSGA-II with a smaller population size can still compute good approximations of the Pareto front. For this, however, a mild modification of the selection mechanism, proposed earlier (Kukkonen and Deb 2006), was needed. Doerr and Qu (2022b) conducted the first mathematical runtime analysis of the NSGA-II on a problem with multimodal objectives. Again, the NSGA-II was found to be as effective as the (G)SEMO algorithms when the population size was chosen suitably.

The three works just discussed consider the NSGA-II as proposed in the original work (Deb et al. 2002) except that they do not employ crossover. The only runtime analysis of the NSGA-II with crossover (prior to this work and the parallel work (Dang et al. 2023)) was conducted by Bian and Qian (2022), namely on the COCZ, ONEMINMAX, and LOTZ benchmarks. The runtime guarantees shown there agree with the ones of Zheng, Liu, and Doerr (2022), so no advantage of crossover was shown. Bian and Qian (2022) also proposed a novel selection mechanism that gives considerable speed-ups.

The question of whether crossover, that is, generating new solutions from two existing ones, is useful or not, is as old as the area of genetic algorithms. Despite its importance, we are still far from having a satisfying answer. It is clear that at all times, the vast majority of evolutionary algorithms used in practice employ crossover. A solid scientific proof for the usefulness of crossover, however, is still missing.

As an example of one of many unsuccessful attempts to explain the power of crossover, we cite the building-block hypothesis (BBH) (Holland 1975), which states that crossover is effective because it allows combining small profitable segments of different solutions. While very convincing on the intuitive level, a simple experimental analysis on a synthetic problem perfectly fulfilling the assumptions of the BBH raised some doubts. In (Mitchell, Forrest, and Holland 1992), a simple randomized hill-climber was found to solve the proposed Royal-Road problem around ten times faster than a comparably simple genetic algorithm using crossover.

Theoretical approaches have been used as well to demonstrate the usefulness of crossover, but again mostly with limited success. The works closest to ours in single-objective optimization regard the JUMP benchmark, which looks like an ideal example for exploiting crossover. Surprisingly, it was much harder than expected to show that crossover is profitable here. The early analyses (Jansen and Wegener 2002; Kötzing, Sudholt, and Theile 2011) managed to show an advantage from crossover only with an unrealistically small crossover probability. Dang et al. (2018) for the first time showed that a standard $(\mu + 1)$ GA with standard parameters optimizes jump functions faster with crossover than without. Other examples proving advantages of crossover include (Sudholt 2005; Lehre and Yao 2011; Doerr, Happ, and Klein 2012; Doerr, Doerr, and Ebel 2015; Dang et al. 2016; Sutton 2021; Antipov, Doerr, and Karavaev 2022), but they appear specific to a particular problem or a particular algorithm.

In multi-objective optimization, so far only three mathematical runtime analyses showing an advantage of crossover exist. Neumann and Theile (2010) proved that crossover is useful when solving multi-criteria versions of the all-pairs-shortest-path (APSP) problem. This work follows the ideas of the corresponding single-objective result (Doerr, Happ, and Klein 2012). Both are somewhat problem-specific in the sense that the formulation of the APSP problem automatically leads to a very strong diversity mechanism, from which crossover substantially profits in both works.

Qian, Yu, and Zhou (2013) showed a substantial advantage from crossover, among others, for the classic COCZ problem. For this, however, a novel initialization is used which, for this problem, results in the initial and all subsequent populations containing both the all-ones and the all-zeroes string. It is clear that this strong diversity greatly helps crossover to become effective.

Both these works regard the SEMO and GSEMO algorithms, two synthetic algorithms proposed for theoretical analyses (Laumanns et al. 2002; Giel 2003). Given this and the particularities of the results, the diversity mechanism implicit in the APSP problem and the particular initialization in Qian, Yu, and Zhou (2013), it is not clear to what extent the insight that crossover is beneficial can be expected to generalize to the broader EMO field.

The third work showing an advantage of crossover in EMO optimization is (Huang et al. 2019). Since it discusses a decomposition-based algorithm, it is very far from our work and we do not detail it further.

Preliminaries

The NSGA-II Algorithm

In the interest of brevity, we only give a brief overview of the algorithm here and refer to (Deb et al. 2002) for a more detailed description. The NSGA-II uses two metrics to completely order any population, which are rank and crowding distance. The ranks are defined recursively based on the dominance relation. All non-dominated individuals have rank 1. Then, given that the individuals of ranks $1, \dots, k$ are defined, the individuals of rank $k+1$ are those not dominated except by individuals of rank k or smaller. This defines a partition of the population into sets F_1, F_2, \dots such that F_i contains all individuals with rank i . Individuals with lower ranks are preferred. The crowding distance, denoted by $cDis(x)$ for an individual x , is used to compare individuals of the same rank. To compute the crowding distances of individuals of rank i with respect to a given objective function f_j , we first sort the individuals in ascending order according to their f_j objective values. The first and last individuals in the sorted list have infinite crowding distance. For the other individuals, their crowding distance is the difference between the objective values of their left and right neighbors in the sorted list, normalized by the difference between the minimum and maximum values. The final crowding distance of an individual is the sum of its crowding distances with respect to each objective function. Among individuals of the same rank, the ones with higher crowding distances are preferred.

The algorithm starts with a random initialization of a parent population of size N . At each iteration, N children are generated from the parent population via a variation operator, and N best individuals among the combined parent and children population survive to the next generation based on their ranks and, as a tie-breaker, the crowding distance (remaining ties are broken randomly). In each iteration, the critical rank i^* is the rank such that if we take all individuals of ranks smaller than i^* , the total number of individuals will be less than or equal to N , but if we also take all individu-

als of rank i^* , the total number of individuals will be over N . Thus, all individuals of rank smaller than i^* survive to the next generation, and for individuals of rank i^* , we take the individuals with the highest crowding distance, breaking ties randomly, so that in total exactly N individuals are kept. In practice, the algorithm is run until some stopping criterion is met. In our mathematical analysis, we are interested in how long it takes until the full Pareto front is covered by the population if the algorithm is not stopped earlier. For that reason, we do not specify a termination criterion.

To create the offspring, the algorithm selects $N/2$ pairs of individuals from the parent population (possibly with repetition). For each pair, with probability 0.9, we generate two intermediate offspring via a 2-offspring uniform crossover (that is, for each position independently, with probability 0.5, the first child inherits the bit from the first parent, and otherwise from the second parent; the bits from the two parents that are not inherited by the first child make up the second child). Bit-wise mutation is then performed on these two intermediate offspring (that is, each bit is flipped independently with probability $\frac{1}{n}$). With the remaining 0.1 probability, this mutation is performed directly on the two parents.

Different methods can be employed to select the parents that are used to create the offspring population (that is, the aforementioned $N/2$ pairs). i) Fair selection: Each individual appears exactly once in a pair, apart from this, the pairing is random. ii) Uniform selection: Each pair consists of two random individuals. iii) N independent binary tournaments: for N times, uniformly at random sample 2 different parents and conduct a binary tournament between the two, i.e., select the one with the lower rank, breaking ties by selecting the one with the larger crowding distance, breaking remaining ties randomly; form $N/2$ pairs from the winners randomly. iv) Two-permutation tournament scheme: Generate two random permutations π_1 and π_2 of P_t and conduct a binary tournament between $\pi_j(2i-1)$ and $\pi_j(2i)$ for all $i \in [1..N/2]$ and $j \in \{1, 2\}$; form a pair from the two winners in each interval of length 4 in a permutation.

The $(\mu + 1)$ Genetic Algorithm

The $(\mu + 1)$ GA maintains a population of μ individuals which are randomly initialized at the beginning of a run. In each generation, a new individual is created. With a constant probability p_c , it is created by selecting two parents from the population uniformly at random, crossing them over, and then applying mutation to the resulting offspring. With probability $1 - p_c$, a single individual is selected and only mutation is applied. At the end of the generation, the worst individual is removed from the population, with ties broken randomly.

Similarly to our analysis of the NSGA-II, in our analysis of the $(\mu + 1)$ GA, we consider applying uniform crossover with probability $p_c = 0.9$. Here crossover only produces one child where each bit has 50% chance coming from the first parent and 50% chance coming from the second. Bit-wise mutation of rate $\frac{1}{n}$ is employed.

Benchmark Problems

For $x \in \{0, 1\}^n$, let $|x|_0$ and $|x|_1$ denote the number of 0-bits and 1-bits in x , respectively. Let $k = \lfloor 2 \cdot n/4 \rfloor$. The function $\text{JUMP}_{n,k} = f : \{0, 1\}^n \rightarrow \mathbb{R}$ was proposed by Droste, Jansen, and Wegener (2002) and serves as the prime example to study how randomized search heuristics cope with local optima (Jansen and Wegener 2002; Jägersküpper and Storch 2007; Doerr et al. 2017; Hasenöhl and Sutton 2018; Lissovoi, Oliveto, and Warwicker 2019; Corus, Oliveto, and Yazdani 2020; Benbaki, Benomar, and Doerr 2021; Hevia Fajardo and Sudholt 2021; Antipov, Buzdalov, and Doerr 2022; Doerr 2022; Rajabi and Witt 2022; Witt 2023). It is defined by

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

for all $x \in \{0, 1\}^n$. The aim is to maximize f . It has a valley of low fitness around its optimum, which can be crossed only by flipping the k correct bits if no solutions of lower fitness are accepted.

The function $\text{ONEJUMPZEROJUMP}_{n,k} = (f_1, f_2) : \{0, 1\}^n \rightarrow \mathbb{R}^2$, proposed by Doerr and Zheng (2021), is defined by

$$f_1(x) = \begin{cases} k + |x|_1, & \text{if } |x|_1 \leq n - k \text{ or } x = 1^n, \\ n - |x|_1, & \text{else;} \end{cases}$$

$$f_2(x) = \begin{cases} k + |x|_0, & \text{if } |x|_0 \leq n - k \text{ or } x = 0^n, \\ n - |x|_0, & \text{else;} \end{cases}$$

for all $x \in \{0, 1\}^n$; here $|x|_0$ is the number of 0-bits in x . The aim is to maximize both f_1 and f_2 , two multimodal objectives. The first objective is the classical $\text{JUMP}_{n,k}$ function. The second objective is isomorphic to the first, with the roles of zeroes and ones exchanged.

According to Theorem 2 of Doerr and Zheng (2021), the Pareto set of the benchmark is $S^* = \{x \in \{0, 1\}^n \mid |x|_1 = [k..n - k] \cup \{0, n\}\}$, and the Pareto front $F^* = f(S^*)$ is $\{(a, 2k + n - a) \mid a \in [2k..n] \cup \{k, n + k\}\}$, making the size of the front $n - 2k + 3$. We define the inner part of the Pareto set by $S_I^* = \{x \mid |x|_1 \in [k..n - k]\}$, and the inner part of the Pareto front by $F_I^* = f(S_I^*) = \{(a, 2k + n - a) \mid a \in [2k..n]\}$.

Doerr and Zheng (2021) proved that the SEMO cannot optimize this benchmark. The GSEMO can and does so in expected time $O(n^{k+1})$. Doerr and Qu (2022b) showed that when using a population of size $N \geq 4(n - 2k + 3)$ to optimize this benchmark, the NSGA-II algorithm never loses a Pareto-optimal solution value once found. Then $O(n^k)$ iterations are needed in expectation to find the full Pareto front (all Pareto optimal solution values).

Runtime Analysis of the NSGA-II

In this section, we analyze¹ the complexity of the crossover-based NSGA-II algorithm with population size $N = c(n -$

¹For reasons of space, some proofs had to be omitted in this extended abstract. They can be found in the preprint (Doerr and Qu 2022a).

$2k + 3)$ for some $c > 4$. We consider all four different ways of selecting the parents for variation described in the previous section. For any generation t of a run of the algorithm, we use P_t to denote the parent population and R_t to denote the combined parent and offspring population. For individuals x and y , we use $H(x, y)$ to denote the Hamming distance between x and y .

Doerr and Qu (2022b) proved that the NSGA-II in an expected time (number of fitness evaluations) of $O(Nn^k)$ covers the entire Pareto front of the ONEJUMPZEROJUMP benchmark. This time mostly stems from the waiting times to generate the extremal points on the front, i.e., the all-zeroes string 0^n and the all-ones string 1^n . Since individuals with lower fitness values are not kept in the population, these two extremal points can only be created from individuals with $i \in [k..n - k]$ 1-bits. Crossing the fitness valley of length k happens for an individual with probability $\Theta(n^{-k})$, resulting in the $O(Nn^k)$ runtime.

To profit from crossover, we exploit the fact that there can be diverse individuals at the outermost points of F_I^* . For example if in a P_t there are two individuals x and y such that $|x|_1 = |y|_1 = n - k$ and x and y are different, i.e., $H(x, y) = 2d$ for $d \geq 1$, crossing them over creates an individual z such that $|z|_1 = n - k + d$ with probability $\Theta(\frac{1}{2^{2d}})$. Then mutating z gives 1^n with probability $\Theta(\frac{1}{n^{k-d}})$. Therefore, the probability that 1^n is generated in iteration t is $\Omega(\frac{1}{2^{2d}n^{k-d}})$, making the waiting time smaller than $O(n^k)$. Following this idea, our analysis focuses on the runtime needed to create, maintain, and in the end, take advantage of such diversity at the outermost points of F_I^* .

First, some observations on the selection and survival probabilities of rank-1 individuals will help us argue that diversity can be created and maintained.

Lemma 1. *Consider an iteration t of the NSGA-II algorithm optimizing the $\text{ONEJUMPZEROJUMP}_{n,k}$ benchmark with population size $N = c(n - 2k + 3)$ for some $c > 4$. With any of the four selection methods, the probability that two rank-1 individuals $x, y \in P_t$ are both selected to mutate in iteration t is $\Theta(1)$.*

Lemma 2. *Consider an iteration t of the NSGA-II algorithm optimizing the $\text{ONEJUMPZEROJUMP}_{n,k}$ benchmark with population size $N = c(n - 2k + 3)$ for some $c > 4$ where $F^* \not\subseteq f(R_t)$. Suppose x and y are two rank-1 individuals in R_t . Then with probability at least $(\frac{c-4}{2c})^2$, we have $x, y \in P_{t+1}$.*

Now, we give a lemma on the runtime needed to create and maintain diversity at the outermost points of F_I^* . The idea is that in an iteration t , through mutation, an individual x with $n - k$ 1-bits can generate an individual y with the same number of 1-bits but different from x (e.g., $H(x, y) = 2$), creating diversity among individuals with $n - k$ 1-bits. Then by Lemma 2, x and y survive to iteration $t + 1$ with a constant probability, where further diversity can be created in the same way. Accumulating diversity in k iterations, we will have two individuals both with $n - k$ 1-bits and having a Hamming distance of $2k$.

Lemma 3. Consider an iteration t of the NSGA-II algorithm optimizing the ONEJUMPZEROJUMP $_{n,k}$ benchmark for $k = o(\sqrt{n})$ with population size $N = c(n - 2k + 3)$ for some $c > 4$. With any of the four parent selection methods, applying uniform crossover with probability 0.9 and bit-wise mutation, if there is $x \in P_t$ such that $|x|_1 = n - k$, then in another $O(\frac{(Kn)^k}{(k-1)!})$ iterations, for $K = (\frac{2c}{c-4})^2$, in expectation, the parent population will contain x and y such that $|x|_1 = |y|_1 = n - k$ and $H(x, y) = 2k$.

Theorem 4. Consider an iteration t of the NSGA-II algorithm optimizing the ONEJUMPZEROJUMP $_{n,k}$ benchmark for $k = o(\sqrt{n})$ with population size $N = c(n - 2k + 3)$ for some $c > 4$. Suppose $1^n \notin P_t$. With any of the four parent selection methods, applying uniform crossover with probability 0.9 and bit-wise mutation, if there is $x \in P_t$ such that $|x|_1 = n - k$, then in expectation the algorithm needs another $O(\frac{N^2(Cn)^k}{(k-1)!})$ fitness evaluations, for $C = (\frac{4c}{c-4})^2$, to find 1^n .

Note that the case for finding 0^n is symmetrical. Therefore, it takes $O(\frac{N(Cn)^k}{(k-1)!})$ iterations to find both 1^n and 0^n . To cover any other point on the front, since there is a 0.1 chance that crossover is not applied and since covered points remain covered, the $O(Nn \log n)$ time for covering the whole inner part of the Pareto front shown in Doerr and Qu (2022b) remains valid (and is a lower order term compared to the time to generate the extremal points). In conclusion, in total, with uniform crossover happening at a constant probability, the number of iterations taken is $O(\frac{N(Cn)^k}{(k-1)!})$.

Our analysis has revealed that the advantage of using crossover comes from the fact that now the algorithm does not need to flip all k bits at once to cross the fitness valley. Still, k bits need to be flipped for an individual x with $n - k$ bits of 1 to create another individual with $n - k$ bits of 1 that is different enough from x so that the two can produce 1^n by only crossover. This explains the n^k term in the runtime proven. However, now these k bits can be flipped one at a time in k iterations, and this happens with probability larger than flipping them all at once since at each iteration the algorithm only needs to flip one bit from multiple available choices. This explains the $\frac{1}{(k-1)!}$ term in the proven runtime. We note that a disadvantage is that flipping one bit at a time means the diversity needed is created in k iterations and in all these k iterations, the diversity that has been created so far needs to be maintained, which is where the C^k term in the runtime comes from. Another disadvantage is that in the end, creating the extremal point through only crossover requires pairing up two particular individuals, which happens with probability $\Theta(1/N)$ only. Therefore, crossover brings an asymptotic speed-up only when $(k-1)!$ outweighs NC^k , i.e., when $k \geq c \log(N) / \log \log(N)$ where c is a suitably large constant. Our experiments, however, show that crossover is profitable already from $k = 2$ on.

Runtime Analysis of the $(\mu + 1)$ GA

Dang et al. (2018) showed that the $(\mu + 1)$ GA with population size $\mu = O(n)$ optimizes the JUMP $_{n,k}$ function,

$k \geq 3$, in expected time $O(\frac{n^k}{\min\{\mu, n/\log n\}})$. Hence for $\mu = \Theta(n/\log n)$, a speed-up from crossover of $\Omega(n/\log n)$ was shown. The proof of this result builds on a complex analysis of the population dynamics. With additional arguments, the requirements on μ were relaxed in (Doerr et al. 2022), but still a logarithmic population size was needed to obtain a speed-up of $\Omega(n)$. In this section, we provide a simpler analysis using the insights from the previous section and prove that crossover is helpful even when μ is $\Theta(1)$.

Dang et al. (2018, Lemma 1) showed that the expected time that the algorithm takes so that the entire population reaches the local optimum, i.e., all individuals have $n - k$ bits of 1, is $O(n\sqrt{k}(\mu \log \mu + \log n))$. We reuse this result, but now show that after the entire population has reached the local optimum, crossover decreases the waiting time to find the all-ones string with the $(\mu + 1)$ GA in a similar fashion as for the NSGA-II. We note that crossover does not interfere with mutation, so as in (Doerr and Qu 2022b) we have an $\Omega(n^{-k})$ chance to generate the extremal points via mutation and thus have an $O(n^k)$ iterations runtime guarantee also when crossover is used with rate 0.9.

Lemma 5. Consider the $(\mu + 1)$ GA optimizing the JUMP $_{n,k}$ function with $k = o(\sqrt{n})$, applying uniform crossover with probability 0.9 and bit-wise mutation. Suppose that in some iteration t the entire population is at the local optimum. Then in another $O((10e\frac{\mu+1}{\mu-1})^k \mu^{k-1} k \frac{n^k}{k!})$ iterations in expectation, the population contains x and y such that $|x|_1 = |y|_1 = n - k$ and $H(x, y) = 2k$.

Theorem 6. Consider the $(\mu + 1)$ GA optimizing the JUMP $_{n,k}$ function, applying uniform crossover with probability 0.9 and bit-wise mutation. The number of iterations needed in expectation is $O((40e\mu(\mu + 1))^k (\frac{1}{\mu-1})^{k-1} \frac{10e}{9} \frac{n^k}{(k-1)!} + n\sqrt{k}(\mu \log \mu + \log n))$.

The runtime proven in Theorem 6 is minimized when $\mu = \Theta(1)$. Take $\mu = 2$ for example. There the speed-up compared with the expected $\Theta(n^k)$ runtime of EAs without crossover is $\Omega(\frac{(k-1)!}{(240e)^k})$, which is a real speed-up for $k = \omega(1)$. In Dang et al. (2018), crossover is only shown to lead to speed up when μ is large. Specifically, for $k = 2$, the best speed-up is observed for $\mu = \Theta(\sqrt{n/\log n})$ and for $k \geq 3$, the best speed-up is observed for $\mu = \Theta(n)$. For $\mu = \Theta(1)$ however, no real speed-up can be inferred from Dang et al. (2018). Our arguments here demonstrate that crossover is helpful for the $(\mu + 1)$ GA even for a small population.

Experiments

To complement our theoretical results, we also experimentally evaluate some runs of the NSGA-II algorithm on the ONEJUMPZEROJUMP benchmark, and the $(\mu + 1)$ GA on the JUMP benchmark.

NSGA-II Optimizing ONEJUMPZEROJUMP

Settings We implemented the algorithm as described in the preliminaries section in Python. We regarded the problem sizes $n = 50$ and $n = 100$ and the jump size $k = 2$.

	$p_c = 0$	$p_c = 0.9$
$n = 50, N = 2(n - 2k + 3)$	247,617	190,577
$n = 50, N = 4(n - 2k + 3)$	416,284	147,921
$n = 100, N = 2(n - 2k + 3)$	2,411,383	1,954,681
$n = 100, N = 4(n - 2k + 3)$	3,858,084	1,322,046

Table 1: Average runtime of the NSGA-II on the ONEJUMPZEROJUMP benchmark with $k = 2$.

We used the population sizes $N = 2(n - 2k + 3)$ and $N = 4(n - 2k + 3)$ and employed fair selection for variation. With probability p_c , uniform crossover is applied followed by bit-wise mutation, otherwise only mutation is performed. We regarded the crossover rates $p_c = 0$ (no crossover) and $p_c = 0.9$. We conducted 10 independent repetitions per setting.

We note that Doerr and Qu (2022b) already present some results on the runtime (for $k = 3$, $N/(2n - k + 3) = 2, 4, 8$, and $n = 20, 30, 40$). Therefore, we chose these settings to present new data points, in particular, for smaller k and larger problem sizes.

Results Table 1 contains the average number of fitness evaluations taken by the NSGA-II algorithm to cover the entire Pareto front of the ONEJUMPZEROJUMP $_{n,k}$ benchmark for $n \in \{50, 100\}$ and $k = 2$. For all results, a standard deviation between 50% to 80% of the mean was observed.

The data clearly supports our claim that crossover speeds up the NSGA-II on the ONEJUMPZEROJUMP benchmark (by a factor of between 1.2 and 2.9). The benefit of crossover is much more pronounced for the larger of the two population sizes. We suspect that this is because when the population becomes larger, there will be more individuals close to the outermost points of F_I^* . As a result, it is easier for the algorithm to reach and maintain diversity there, which as our analysis has suggested, is what makes crossover beneficial.

The runtimes for $k = 3$ (and smaller problem sizes) reported by Doerr and Qu (2022b) showed speed-ups from crossover between a factor of 3.9 and 9. This fits our mathematical analysis which showed more significant speed-ups for larger problem difficulties k .

To support our reasoning on how efficient crossover could be for different population sizes, we have also recorded for $n = 50$ the diversity among the individuals with k and $n - k$ bits of 1 throughout the runs of the experiments. Specifically, for every $n^k/50$ iterations, we look at the individuals in the parent population with k and $n - k$ bits of 1. Among all individuals with k bits of 1, we calculate the Hamming distances between any two individuals, and record the maximum distance divided by 2. This number gives us an idea of how much a lucky crossover can decrease the waiting time of finding 0^n , since as discussed in our analysis, pairing up two individuals with k bits of 1 whose hamming distance is $2d$ means there is an $\Omega(\frac{1}{2^{2d}n^{k-d}})$ chance of creating the all-zeroes string. We do the same for the individuals with $n - k$ bits of 1. Note that the greatest this number can be is k . Since as shown in Doerr and Qu (2022b), the runtime is dominated by the waiting time needed to find the all-ones

and the all-zeroes strings after the inner part of the Pareto front has been covered, we are mostly interested in how the diversity develops in that phase. To this end, we discard data points recorded when the inner part of the Pareto front has not been fully covered, and those recorded after one of the extremal points has already been discovered. For one run of the experiment, we average the data points recorded for k and $n - k$ bits of 1 together in the end. Finally we average the mean values obtained from the 10 repetitions. For $N = 2(n - 2k + 3)$, we have observed that the diversity measure is $0.76_{\pm 0.62}$, while for $N = 4(n - 2k + 3)$, it is $0.99_{\pm 0.45}$. This means indeed there is more diversity on the boundaries with a larger population, explaining why with $N = 4(n - 2k + 3)$ the speed-up from crossover is more than that with $N = 2(n - 2k + 3)$. We note that this does not mean that large population sizes are preferable since the higher cost of one iteration has to be taken into account. Due to the high runtimes, we do not have yet data allowing us to answer this question conclusively, but preliminary experiments for $n = 50$ suggest that a population size of $N = 8(n - 2k + 3)$ already gives inferior results.

We note that the speed-up from crossover empirically is more profound than what we have shown theoretically. Though it is normal that mathematical runtime guarantees cannot capture the full strength of an algorithm, as a first attempt to understand where the discrepancy comes from we also recorded for each run how the all-ones string is generated (for the case $n = 100$ and $N = 2(n - 2k + 3)$). As a result, among the 10 runs that we have done, 9 times crossover has participated in the generation of 1^n . Out of these 9 times, there are 5 times where crossover by itself has created 1^n . However, for all of those 5 times, only one of the parents has $n - k = 98$ 1-bits, while the other parent has between 95 to 97 1-bits. Among all the runs, we have observed only once that both of the parents have 98 bits of 1. This suggests that crossover also profits from diversity between individuals of different objective values, a fact not exploited in our runtime analysis.

The $(\mu + 1)$ GA Optimizing JUMP

Settings We implemented the algorithm as described in the preliminaries section in Python. We use the problem sizes $n = 100$ and $n = 1000$, a jump size of $k = 4$, the population sizes $\mu = 2^i$ for $i = [0..9]$ for $n = 100$, and $\mu = 2^i$ for $i = [5..11]$ for $n = 1000$. With 0.9 chance uniform crossover is applied followed by bit-wise mutation, otherwise only bit-wise mutation is performed. Each setting was run independently 10 times.

Runtime Figure 1 gives the average number of fitness evaluations needed for the $(\mu + 1)$ GA to optimize JUMP $_{n,k}$ for $n \in \{100, 1000\}$. The results confirm that crossover is beneficial for the $(\mu + 1)$ GA optimizing JUMP as predicted from the runtime analyses in (Dang et al. 2018) and this work. The experimental results suggest that the best speed-up is observed for large populations (128 for $n = 100$ and 512 for $n = 1000$), different from our analysis. However, already smaller populations give significant speed-ups. For example, when $n = 100$ and $k = 4$, merely with

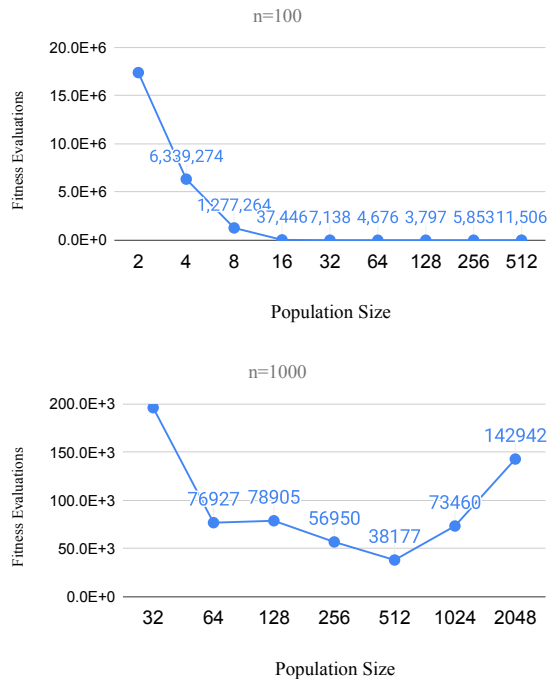


Figure 1: Average runtime of the $(\mu + 1)$ GA to optimize $\text{JUMP}_{n,k}$, $k = 4$, using uniform crossover with rate 0.9.

$\mu = 2$, crossover is able to decrease the runtime from over $((1/n)^k(1 - 1/n)^{n-k})^{-1} \geq 10^8$, the expected waiting time for a successful mutation, to 1.7×10^7 . With $\mu = 4$, the runtime is further decreased to 6.3×10^6 . These results suggest that the effects exploited in our analysis contribute measurably significantly to the advantages of crossover, in particular, when $\mu = \Theta(1)$ where the analysis in Dang et al. (2018) could not show a speed-up.

Conclusions and Future Works

In this work, we conducted a mathematical runtime analysis of the NSGA-II that shows a speed-up from crossover (on the ONEJUMPZEROJUMP benchmark), already for small population sizes (larger than four times the Pareto front, a number required already for previous runtime analyses of a mutation-only version of the NSGA-II). Interestingly, the proven gain from crossover increases with the difficulty parameter k of the ONEJUMPZEROJUMP benchmark. Our results are very different from previous runtime analyses of crossover-based algorithms. Hence our work has detected a novel way how crossover-based algorithms can leave local optima. As a side result, we show that our arguments can be employed for the $(\mu + 1)$ GA, showing significant speed-ups there as well, again from small (constant) population sizes on and increasing significantly with the difficulty parameter k .

Our experimental results, similar to the ones in (Dang et al. 2018), confirm the proven advantages of crossover, but also show that crossover is much more powerful than what

the mathematical proofs could distill. For the NSGA-II, for example, we observe considerable speed-ups already for the smallest possible value $k = 2$ of the difficulty parameter. Trying to explain these via proven results appears as an interesting problem for future research.

As a second direction for further research, we note that we did not prove any lower bounds, so we have no estimate on how far our runtime guarantees are from the truth. So far lower bounds have only been proven for the mutation-based NSGA-II (Doerr and Qu 2023). Understanding the population dynamics of a crossover-based algorithm well enough to admit reasonable lower bounds surely needs significantly stronger arguments since now it is not sufficient to understand how many individuals with a certain objective value exist in the population, but also their genotype is important.

Acknowledgments

This work was supported by a public grant as part of the Investissements d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH.

References

- Antipov, D.; Buzdalov, M.; and Doerr, B. 2022. Fast mutation in crossover-based algorithms. *Algorithmica*, 84: 1724–1761.
- Antipov, D.; Doerr, B.; and Karavaev, V. 2022. A rigorous runtime analysis of the $(1 + (\lambda, \lambda))$ GA on Jump functions. *Algorithmica*, 84: 1573–1602.
- Auger, A.; and Doerr, B., eds. 2011. *Theory of Randomized Search Heuristics*. World Scientific Publishing.
- Bäck, T. 1993. Optimal mutation rates in genetic search. In *International Conference on Genetic Algorithms, ICGA 1993*, 2–8. Morgan Kaufmann.
- Benbaki, R.; Benomar, Z.; and Doerr, B. 2021. A rigorous runtime analysis of the 2-MMAS_{ib} on jump functions: ant colony optimizers can cope well with local optima. In *Genetic and Evolutionary Computation Conference, GECCO 2021*, 4–13. ACM.
- Bian, C.; and Qian, C. 2022. 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*, 428–441. Springer.
- Corus, D.; Oliveto, P. S.; and Yazdani, D. 2020. When hypermutations and ageing enable artificial immune systems to outperform evolutionary algorithms. *Theoretical Computer Science*, 832: 166–185.
- Dang, D.; Friedrich, T.; Kötzing, T.; Krejca, M. S.; Lehre, P. K.; Oliveto, P. S.; Sudholt, D.; and Sutton, A. M. 2016. Escaping local optima with diversity mechanisms and crossover. In *Genetic and Evolutionary Computation Conference, GECCO 2016*, 645–652. ACM.
- Dang, D.; Friedrich, T.; Kötzing, T.; Krejca, M. S.; Lehre, P. K.; Oliveto, P. S.; Sudholt, D.; and Sutton, A. M. 2018. Escaping local optima using crossover with emergent diversity. *IEEE Transactions on Evolutionary Computation*, 22: 484–497.

- Dang, D.-C.; Opris, A.; Salehi, B.; and Sudholt, D. 2023. A proof that using crossover can guarantee exponential speed-ups in evolutionary multi-objective optimisation. In *Conference on Artificial Intelligence, AAAI 2023*. AAAI Press.
- Deb, K.; Pratap, A.; Agarwal, S.; and Meyarivan, T. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6: 182–197.
- Doerr, B. 2022. Does Comma Selection Help to Cope with Local Optima? *Algorithmica*, 84: 1659–1693.
- Doerr, B.; Doerr, C.; and Ebel, F. 2015. From black-box complexity to designing new genetic algorithms. *Theoretical Computer Science*, 567: 87–104.
- Doerr, B.; Echarchaoui, A.; Jamal, M.; and Krejca, M. S. 2022. Lasting Diversity and Superior Runtime Guarantees for the $(\mu + 1)$ Genetic Algorithm. *CoRR*, abs/2302.12570.
- Doerr, B.; Happ, E.; and Klein, C. 2012. Crossover can provably be useful in evolutionary computation. *Theoretical Computer Science*, 425: 17–33.
- Doerr, B.; Le, H. P.; Makhmara, R.; and Nguyen, T. D. 2017. Fast genetic algorithms. In *Genetic and Evolutionary Computation Conference, GECCO 2017*, 777–784. ACM.
- Doerr, B.; and Neumann, F., eds. 2020. *Theory of Evolutionary Computation—Recent Developments in Discrete Optimization*. Springer. Also available at http://www.lix.polytechnique.fr/Labo/Benjamin.Doerr/doerr_neumann_book.html.
- Doerr, B.; and Qu, Z. 2022a. The First Mathematical Proof That Crossover Gives Super-Constant Performance Gains For the NSGA-II. *CoRR*, abs/2208.08759.
- Doerr, B.; and Qu, Z. 2022b. A first runtime analysis of the NSGA-II on a multimodal problem. In *Parallel Problem Solving From Nature, PPSN 2022*, 399–412. Springer.
- Doerr, B.; and Qu, Z. 2023. From Understanding the Population Dynamics of the NSGA-II to the First Proven Lower Bounds. In *Conference on Artificial Intelligence, AAAI 2023*. AAAI Press.
- Doerr, B.; and Wietheger, S. 2022. A Mathematical Runtime Analysis of the Non-dominated Sorting Genetic Algorithm III (NSGA-III). *CoRR*, abs/2211.08202.
- Doerr, B.; and Zheng, W. 2021. Theoretical analyses of multi-objective evolutionary algorithms on multi-modal objectives. In *Conference on Artificial Intelligence, AAAI 2021*, 12293–12301. AAAI Press.
- Droste, S.; Jansen, T.; and Wegener, I. 2002. On the analysis of the $(1+1)$ evolutionary algorithm. *Theoretical Computer Science*, 276: 51–81.
- Giel, O. 2003. Expected runtimes of a simple multi-objective evolutionary algorithm. In *Congress on Evolutionary Computation, CEC 2003*, 1918–1925. IEEE.
- Gutjahr, W. J. 2008. First steps to the runtime complexity analysis of ant colony optimization. *Computers & Operations Research*, 35: 2711–2727.
- Hasenöhr, V.; and Sutton, A. M. 2018. On the runtime dynamics of the compact genetic algorithm on jump functions. In *Genetic and Evolutionary Computation Conference, GECCO 2018*, 967–974. ACM.
- Hevia Fajardo, M. A.; and Sudholt, D. 2021. Self-adjusting offspring population sizes outperform fixed parameters on the cliff function. In *Foundations of Genetic Algorithms, FOGA 2021*, 5:1–5:15. ACM.
- Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Huang, Z.; Zhou, Y.; Chen, Z.; and He, X. 2019. Running time analysis of MOEA/D with crossover on discrete optimization problem. In *Conference on Artificial Intelligence, AAAI 2019*, 2296–2303. AAAI Press.
- Jägersküpper, J.; and Storch, T. 2007. When the plus strategy outperforms the comma strategy and when not. In *Foundations of Computational Intelligence, FOCI 2007*, 25–32. IEEE.
- Jansen, T. 2013. *Analyzing Evolutionary Algorithms – The Computer Science Perspective*. Springer.
- Jansen, T.; and Wegener, I. 2002. The analysis of evolutionary algorithms – a proof that crossover really can help. *Algorithmica*, 34: 47–66.
- Kötzing, T.; Sudholt, D.; and Theile, M. 2011. How crossover helps in pseudo-Boolean optimization. In *Genetic and Evolutionary Computation Conference, GECCO 2011*, 989–996. ACM.
- Kukkonen, S.; and Deb, K. 2006. Improved pruning of non-dominated solutions based on crowding distance for bi-objective optimization problems. In *Conference on Evolutionary Computation, CEC 2006*, 1179–1186. IEEE.
- Laumanns, M.; Thiele, L.; Zitzler, E.; Welzl, E.; and Deb, K. 2002. Running time analysis of multi-objective evolutionary algorithms on a simple discrete optimization problem. In *Parallel Problem Solving from Nature, PPSN 2002*, 44–53. Springer.
- Lehre, P. K.; and Yao, X. 2011. Crossover can be constructive when computing unique input-output sequences. *Soft Computing*, 15: 1675–1687.
- Lissovoi, A.; and Oliveto, P. S. 2018. On the time and space complexity of genetic programming for evolving boolean conjunctions. In *Conference on Artificial Intelligence, AAAI 2018*, 1363–1370. AAAI Press.
- Lissovoi, A.; Oliveto, P. S.; and Warwicker, J. A. 2019. On the time complexity of algorithm selection hyper-heuristics for multimodal optimisation. In *Conference on Artificial Intelligence, AAAI 2019*, 2322–2329. AAAI Press.
- Mitchell, M.; Forrest, S.; and Holland, J. H. 1992. The Royal Road for Genetic Algorithms: Fitness Landscapes and GA Performance. In *Proc. of the First European Conference on Artificial Life*, 245–254. MIT Press.
- Neumann, F.; and Theile, M. 2010. How crossover speeds up evolutionary algorithms for the multi-criteria all-pairs-shortest-path problem. In *Parallel Problem Solving from Nature, PPSN 2010, Part I*, 667–676. Springer.
- Neumann, F.; and Witt, C. 2010. *Bioinspired Computation in Combinatorial Optimization – Algorithms and Their Computational Complexity*. Springer.
- Neumann, F.; and Witt, C. 2015. On the runtime of randomized local search and simple evolutionary algorithms for

- dynamic makespan scheduling. In *International Joint Conference on Artificial Intelligence, IJCAI 2015*, 3742–3748. AAAI Press.
- Qian, C.; Yu, Y.; and Zhou, Z. 2013. An analysis on recombination in multi-objective evolutionary optimization. *Artificial Intelligence*, 204: 99–119.
- Rajabi, A.; and Witt, C. 2022. Self-adjusting evolutionary algorithms for multimodal optimization. *Algorithmica*, 84: 1694–1723.
- Sasaki, G. H.; and Hajek, B. 1988. The time complexity of maximum matching by simulated annealing. *Journal of the ACM*, 35: 387–403.
- Sudholt, D. 2005. Crossover is provably essential for the Ising model on trees. In *Genetic and Evolutionary Computation Conference, GECCO 2005*, 1161–1167. ACM.
- Sutton, A. M. 2021. Fixed-parameter tractability of crossover: steady-state GAs on the closest string problem. *Algorithmica*, 83: 1138–1163.
- Thierens, D. 2003. Convergence time analysis for the multi-objective counting ones problem. In *Evolutionary Multi-Criterion Optimization, EMO 2003*, 355–364. Springer.
- Witt, C. 2023. How majority-vote crossover and estimation-of-distribution algorithms cope with fitness valleys. *Theoretical Computer Science*, 940: 18–42.
- Zheng, W.; and Doerr, B. 2022a. Better approximation guarantees for the NSGA-II by using the current crowding distance. In *Genetic and Evolutionary Computation Conference, GECCO 2022*, 611–619. ACM.
- Zheng, W.; and Doerr, B. 2022b. Runtime Analysis for the NSGA-II: Proving, Quantifying, and Explaining the Inefficiency for Three or More Objectives. *CoRR*, abs/2211.13084.
- Zheng, W.; Liu, Y.; and Doerr, B. 2022. A first mathematical runtime analysis of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II). In *Conference on Artificial Intelligence, AAAI 2022*, 10408–10416. AAAI Press.