
Runtime Analysis of Restricted Tournament Selection for Bimodal Optimisation

Edgar Covantes Osuna

edgar.covantes@tec.mx

School of Engineering and Sciences, Tecnológico de Monterrey, Monterrey, 64849, México.

Dirk Sudholt

dirk.sudholt@uni-passau.de

Department of Computer Science, University of Sheffield, United Kingdom and
Chair of Algorithms for Intelligent Systems, University of Passau, Germany.

Abstract

Niching methods have been developed to maintain the population diversity, to investigate many peaks in parallel and to reduce the effect of genetic drift. We present the first rigorous runtime analyses of restricted tournament selection (RTS), embedded in a $(\mu+1)$ EA, and analyse its effectiveness at finding both optima of the bimodal function TWOMAX. In RTS, an offspring competes against the closest individual, with respect to some distance measure, amongst w (window size) population members (chosen uniformly at random with replacement), to encourage competition within the same niche. We prove that RTS finds both optima on TWOMAX efficiently if the window size w is large enough. However, if w is too small, RTS fails to find both optima even in exponential time, with high probability. We further consider a variant of RTS selecting individuals for the tournament *without* replacement. It yields a more diverse tournament and is more effective at preventing one niche from taking over the other. However, this comes at the expense of a slower progress towards optima when a niche collapses to a single individual. Our theoretical results are accompanied by experimental studies that shed light on parameters not covered by the theoretical results and support a conjectured lower runtime bound.

Keywords

Niching methods, runtime analysis, restricted tournament selection, evolutionary algorithm.

1 Introduction

One of the major challenges when using Evolutionary Algorithms (EAs) is to maintain the diversity in the population in order to prevent premature convergence. One way of maintaining diversity is to use niching methods, which are based on the mechanics of natural ecosystems (Shir, 2012). Niches can be viewed as subspaces in the environment that can support different types of life. A specie is defined as a group of individuals with similar features, capable of interbreeding among themselves, but unable to breed with individuals outside their group. Species can be defined as similar individuals of a specific niche in terms of similarity metrics. In evolutionary algorithms the term niche is used for the search space domain, and species for the set of individuals with similar characteristics.

Niching methods have been developed to reduce the effect of genetic drift resulting from the selection operator in standard EAs, to maintain the population diversity,

E. Covantes Osuna and D. Sudholt

and to allow the EA to investigate many peaks simultaneously, thus avoiding getting trapped in local optima before the fitness landscape is explored properly (Sareni and Krahenbuhl, 1998). This is often done by modifying the selection process of individuals, taking into account not only the value of the fitness function but also the distribution of individuals in the space of genotypes or phenotypes (Glibovets and Gulayeva, 2013). Many niching techniques have been introduced to solve problems where it is necessary to identify multiple optima (multimodal problems), either local or global optima. The main goal for these techniques is to form and maintain multiple, diverse, final solutions for an exponential to infinite time period with respect to population size, whether these solutions are of identical fitness or of varying fitness (Shir, 2012; Črepinšek et al., 2013; Glibovets and Gulayeva, 2013; Squillero and Tonda, 2016). Given such a variety of mechanisms to choose from, it is often not clear which mechanism is the best choice for a particular problem.

Most of the analyses and comparisons made between niching methods used empirical investigations on benchmark functions (Sareni and Krahenbuhl, 1998; Singh and Deb, 2006). Theoretical runtime analyses have been performed that rigorously quantify the expected time needed to find one or several global optima (Friedrich et al., 2009; Oliveto et al., 2019; Covantes Osuna and Sudholt, 2019, 2020). One example where theoretical results are used to inform the choice of the EA parameters' values in an empirical performance study of EAs with diversity mechanisms can be found in Covantes Osuna and Sudholt (2018a). Both approaches are important to understand how these mechanisms impact the performance of EAs and whether and how they enhance the search for good individuals. These different approaches can help to explain when a niching mechanism should be used, which niching mechanism works best, and how to set parameters.

Previous theoretical studies (Friedrich et al., 2009; Oliveto et al., 2019; Covantes Osuna and Sudholt, 2019, 2020) compared the expected running time of different diversity mechanisms when embedded in a simple baseline EA, the $(\mu+1)$ EA. All mechanisms were considered on the well-known bimodal function $TWOMAX(x) := \max\{n - \sum_{i=1}^n x_i, \sum_{i=1}^n x_i\}$. $TWOMAX$ consists of two different symmetric slopes (or branches) $ZEROMAX$ and $ONEMAX$ with 0^n and 1^n as global optima, respectively, and the goal is to evolve a population that contains both optima¹.

$TWOMAX$ was chosen because it is simply structured, hence facilitating a theoretical analysis, and it is hard for EAs to find both optima as they have the maximum possible Hamming distance. The results allowed for a fair comparison across a wide range of diversity mechanisms, revealing that some mechanisms like fitness diversity or avoiding genotype duplicates and probabilistic crowding perform badly, while other mechanisms like fitness sharing, clearing or deterministic crowding perform surprisingly well (see Table 1 and Section 2).

We contribute to this line of work by studying the performance of the crowding mechanism called *restricted tournament selection* (RTS). This mechanism is a well-known technique as covered in tutorials and surveys for diversity-preserving mechanisms (Shir, 2012; Črepinšek et al., 2013; Glibovets and Gulayeva, 2013; Squillero and

¹In Friedrich et al. (2009) an additional fitness value for 1^n was added to distinguish between a local optimum 0^n and a unique global optimum. There the goal was to find the global optimum, and all approaches had a baseline probability of $1/2$ of climbing up the right branch by chance. We use the same approach as Oliveto et al. (2019); Covantes Osuna and Sudholt (2019), and consider the original definition of $TWOMAX$ and the goal of finding both global optima. The discussion and presentation of previous work from Friedrich et al. (2009) is adapted to our setting. We refer to Sudholt (2020) for details.

Tonda, 2016) and compared in empirical investigations (Sareni and Krahenbuhl, 1998; Singh and Deb, 2006; Covantes Osuna and Sudholt, 2018a).

Restricted tournament selection is a modification of the classical tournament selection for multimodal optimisation that exhibits niching capabilities. For each offspring, RTS sets up a tournament of w (window size) individuals, chosen uniformly at random with replacement² from the population. The offspring competes against the closest individual with respect to some distance measure from the tournament and the best individual is selected for the next generation. This form of tournament restricts an entering individual from competing with others too different from it (Harik, 1995). RTS has been analysed empirically for the classical comparison between crowding mechanisms for multimodal optimisation as a replacement strategy (Qu and Suganthan, 2010; Sareni and Krahenbuhl, 1998; Singh and Deb, 2006). Recent applications for engineering problems with multimodal domains include facility layout design (García-Hernández et al., 2015) and the design of product lines (Tsafarakis, 2016) with reported better results compared to the other variants without RTS.

However, we are lacking a good understanding of when and why it performs well and how it compares to diversity mechanisms analysed previously. Our contribution is to provide a rigorous theoretical runtime analysis accompanied by experimental studies for this mechanism in the context of the $(\mu+1)$ EA on TWOMAX, to rigorously assess its performance in comparison to other diversity mechanisms. In addition, our goal is to provide insights into the working principles of this mechanism to narrow the gap between theory and practice, and to enhance our understanding of its strengths and weaknesses.

1.1 Our contribution

For the $(\mu+1)$ EA with RTS, we show in Section 3 that the mechanism succeeds in finding both optima of TWOMAX in the same way as deterministic crowding, provided that the window size w is chosen large enough, in time $O(\mu n \log n)$ with high probability (Section 3.1). We also show that, if the window size is too small, then it cannot prevent one branch taking over the other, leading to exponential running times with high probability (Section 3.2).

We further consider a variant of RTS, where the tournament chooses individuals *without replacement*. This simple change tends to make the tournament more diverse. We investigate its effect on the performance of the $(\mu+1)$ EA in Section 4. We show that the positive result for RTS with replacement holds for its variant without replacement for a weaker condition $w \geq \mu$ (Section 4.1). For small window sizes, the situation is less clear as the previous analysis for RTS with replacement breaks down. We show that RTS without replacement will find both optima with certainty once two subpopulations have evolved whose best fitness is above a certain threshold value. This means that takeover cannot happen during the final stages of a run where both branches are being explored. However, during early stages of a run, takeover can still happen under rare conditions.

We also find that, surprisingly, in runs where both optima are found, RTS without replacement seems to take significantly more time. This is because typically a popula-

²We believe that the original definition of RTS picks individuals *with* replacement, that is, it is possible to select multiple copies of one individual, ending up with fewer than w different genotypes. The exact implementation was not explained in Harik (1995), however most calculations in Harik (1995) assume selection with replacement. Oddly enough, from all cited papers here, where RTS has been analysed or used, only García-Martínez et al. (2012) make explicit mention of the selection policy used to select the w individuals (uniformly at random with replacement). We consider a variant *without* replacement in Section 4.

E. Covantes Osuna and D. Sudholt

tion will evolve until only a single individual remains on one branch, and this individual takes time $\Theta((\mu^2/w) \cdot n \log n)$ to evolve a global optimum, under certain assumptions (see Section 4.2). This is by a factor of order μ/w larger than the upper bound of $O(\mu n \log n)$ for RTS with replacement from Section 3.1.

Our theoretical results are accompanied by experimental studies that match the theoretical results and also shed light on parameters not covered by the theoretical results. We performed experiments for RTS with replacement and its variant without replacement in order to observe different aspects of the algorithms such as their ability to find both optima on TWOMAX (Sections 5.1 and 5.2). We further support the conjectured lower bound of $\Omega((\mu^2/w) \cdot n \log n)$ for small values of w by measuring the time needed for RTS without replacement to find both optima on TWOMAX (Section 5.2.1). Finally, we assess when takeover is more likely to happen and which variant is more resilient to takeover (Section 5.3).

This article significantly extends a preliminary conference paper (Covantes Osuna and Sudholt, 2018b) that contained preliminary theoretical results for the original RTS (selecting the tournament with replacement) and preliminary experimental results. In this manuscript, the negative result for RTS with small w in Covantes Osuna and Sudholt (2018b) has been improved.

2 Previous Work and Preliminaries

There has been a line of work comparing various diversity mechanisms on TWOMAX in the context of the simple $(\mu+1)$ EA (see Algorithm 1). The $(\mu+1)$ EA starts with a population of size μ created uniformly at random and generates one offspring via mutation; the resulting offspring competes with an individual selected uniformly at random from the subpopulation with worst fitness and the best individual replaces the worst (in case of ties, the offspring is preferred). Table 1 summarises all known results, including our contributions (shown in bold) and conditions involving population size μ and specific parameters of each diversity mechanism explained below. Results from Friedrich et al. (2009) are adapted to our definition of TWOMAX; see Sudholt (2020) for details.

Algorithm 1 $(\mu+1)$ EA

- 1: Initialise P with μ individuals chosen uniformly at random
 - 2: **while** optimum **not found** **do**
 - 3: Choose $x \in P$ uniformly at random
 - 4: Create y by flipping each bit in x independently with probability $1/n$.
 - 5: Choose $z \in P$ uniformly at random from all individuals with worst fitness in P .
 - 6: **if** $f(y) \geq f(z)$ **then** $P = P \setminus \{z\} \cup \{y\}$ **end if**
 - 7: **end while**
-

The notion of success in previously studied mechanisms was being able to find both optima in (expected) time $O(\mu n \log n)$. To put this time bound in context, it is easy to see that the simple $(1+1)$ EA finds one optimum of TWOMAX in expected time $\Theta(n \log n)$. The time bound $O(\mu n \log n)$ includes an additional factor of μ to account for the overhead of evolving μ individuals instead of one. This overhead is necessary for mechanisms like deterministic crowding as used in Friedrich et al. (2009), which essentially evolves μ independent lineages of $(1+1)$ EAs (Friedrich et al., 2009).

Table 1 shows that not all mechanisms succeed in finding both optima on TWOMAX in (expected) time $O(\mu n \log n)$. Friedrich et al. (2009) showed that the plain $(\mu+1)$ EA and the simple mechanisms like *avoiding genotype* or *fitness duplicates* are not

Table 1: Overview of runtime analyses for the $(\mu+1)$ EA with diversity mechanisms on TWOMAX, showing the probability of finding both optima within (expected) time $O(\mu n \log n)$. Results derived in this paper are shown in bold.

Diversity Mechanism	Success prob.	Conditions
Plain $(\mu+1)$ EA ¹	$o(1)$	$\mu = o(n/\log n)$
No Duplicates ¹		
Genotype	$o(1)$	$\mu = o(\sqrt{n})$
Fitness	$o(1)$	$\mu = \text{poly}(n)$
Deterministic Crowding ¹	$1 - 2^{-\mu+1}$	all μ
Fitness Sharing ($\sigma = n/2$)		
Population-based ¹	1	$\mu \geq 2$
Individual-based ²	1	$\mu \geq 3$
Clearing ($\sigma = n/2$) ³	1	$\mu \geq \kappa n^2$
Probabilistic Crowding ⁴	$2^{-\Omega(n)}$	all μ
Probabilistic Crowding with Scaling ⁴		
General bases α	$2^{-\Omega(n/\alpha)}$	all $\alpha \geq 1$
Very large α	$1 - 2^{-\mu+1}$	$\alpha \geq (1+\Omega(1))en$
Generalised Crowding ⁴		
General scaling factors ϕ	$2^{-\Omega(\phi n)}$	all $\phi \leq 1$
Very small ϕ	$1 - 2^{-\mu+1}$	$\phi \leq \frac{1-\Omega(1)}{e^2 n}$
RTS with replacement		
Small window size w (Theorem 5)	$o(1)$	$\mu = o(n^{1/w})$
Large window size w (Theorem 3)	$1 - 2^{-\mu'+3}$	$w \geq 2.5\mu \ln n$
RTS without replacement		
Large window size w (Theorem 6)	$1 - 2^{-\mu'+3}$	$w \geq \mu$

¹ Friedrich et al. (2009).
² Oliveto et al. (2019).
³ Covantes Osuna and Sudholt (2019).
⁴ Covantes Osuna and Sudholt (2020).

able to prevent the extinction of one branch, ending with the population converging to one optimum, with high probability. *Deterministic crowding* with a sufficiently large population is able to reach both optima with probability $1 - 2^{-\mu+1}$ in expected time $O(\mu n \log n)$. This probability converges to 1 exponentially fast in μ ; for instance, a small population size of $\mu = 10$ already gives a success probability of ≈ 0.998 and for $\mu = 30$ it grows to ≈ 0.9999999981 . A *population-based fitness sharing* approach, constructing the best possible new population amongst parents and offspring, with $\mu \geq 2$ and a sharing radius of $\sigma = n/2$ is able to find both optima in expected runtime $O(\mu n \log n)$. The drawback of this approach is that all possible size μ subsets of this union of size $\mu + \lambda$ (where λ is the offspring population size) need to be examined. This is prohibitive for large μ and λ .

Oliveto et al. (2019) studied the original *fitness sharing* approach and showed that a population size $\mu = 2$ is not sufficient to find both optima in polynomial time; the success probability is only $1/2 - \Omega(1)$. However, with $\mu \geq 3$ fitness sharing again finds both

E. Covantes Osuna and D. Sudholt

optima in expected time $O(\mu n \log n)$. Covantes Osuna and Sudholt (2019) analysed the *clearing* mechanism and showed that it can optimise all functions of unitation—function defined over the number of 1-bits contained in a string—in expected time $O(\mu n \log n)$ when the distance function and parameters like the clearing radius σ , the niche capacity κ (how many winners a niche can support) and μ are chosen appropriately. In the case of large niches, that is, with a clearing radius of $\sigma = n/2$, it is able to find both optima in expected time $O(\mu n \log n)$.

Finally, Covantes Osuna and Sudholt (2020) analysed *probabilistic crowding* (Mengshoel and Goldberg, 1999) (extending preliminary results from Covantes Osuna and Sudholt (2018b), which are not included in this manuscript) and showed that it requires exponential time with overwhelming probability on TWOMAX and general classes of functions with bounded gradients. They showed that probabilistic crowding is unable to evolve solutions that are significantly closer to any global optimum than those found by random search, even when given exponential time $2^{\Omega(n)}$ for all population sizes. Probabilistic crowding requires exponential time even when applying exponential scaling to TWOMAX: for every constant base α , and even values up to $\alpha = O(n^{1-\varepsilon})$, on the function $\alpha^{\text{TWOMAX}(x)}$, the selection pressure is still too low, leading to an exponential time $2^{\Omega(n/\alpha)}$ with overwhelming probability. Only when $\alpha = \Omega(n)$, the selection pressure becomes large enough to enable hill climbing. In this case, probabilistic crowding with scaling is as successful on TWOMAX as deterministic crowding.

Generalised crowding (Galán and Mengshoel, 2010) is a variant that generalises both deterministic and probabilistic crowding through the choice of a parameter called *scaling factor* $\phi \in [0, 1]$ that diminishes the impact of the inferior search point. With $\phi = 1$ we have probabilistic crowding and $\phi = 0$ yields deterministic crowding. Covantes Osuna and Sudholt (2020) showed that if $\phi = \Omega(n^{-1+\varepsilon})$, for a constant $\varepsilon > 0$, then this gives exponential time with overwhelming probability on all functions with bounded gradients. But if $\phi = O(1/n)$, generalised crowding behaves similarly to deterministic crowding and becomes effective on TWOMAX.

The above works did not consider crossover as recombining individuals from different branches is likely to create poor offspring. We therefore consider a $(\mu+1)$ EA using mutation only and we introduce RTS into the definition of the $(\mu+1)$ EA. In RTS a new offspring competes with the closest element with respect to some distance measure from w (*window size*) members selected uniformly at random, with replacement, from the population, and the better individual from this competition is selected. The $(\mu+1)$ EA with RTS (Algorithm 2) is defined similarly as the plain $(\mu+1)$ EA (see Algorithm 1), the $(\mu+1)$ EA with deterministic crowding in Friedrich et al. (2009) and the $(\mu+1)$ EA with probabilistic crowding in Covantes Osuna and Sudholt (2018b) to facilitate comparisons between all the algorithms and all available results.

In Algorithm 2 an individual x is selected uniformly at random as a parent and a new individual y is created in the mutation step. Since we are not considering crossover and only one individual is created, w individuals are selected uniformly at random with replacement and stored in a temporary population P^* . Then in Line 6 an individual z is selected from population P^* with the minimum distance from y (ties are broken uniformly at random), and if the individual y has a fitness at least as good as z , y replaces z .

As distance functions $d(\cdot, \cdot)$ we consider genotypic or Hamming distance, defined as the number of bits that have different values in x and y : $d(x, y) := H(x, y) := \sum_{i=0}^{n-1} |x_i - y_i|$, and phenotypic distances as in Friedrich et al. (2009); Oliveto et al. (2019);

Algorithm 2 ($\mu+1$) EA with restricted tournament selection

- 1: Initialise P with μ individuals chosen uniformly at random
 - 2: **while** stopping criterion **not met do**
 - 3: Choose $x \in P$ uniformly at random
 - 4: Create y by flipping bits in x independently with probability $1/n$.
 - 5: Select w individuals uniformly at random, with replacement, from P and store them in P^* .
 - 6: Choose $z \in P^*$ with the minimum distance to y .
 - 7: **if** $f(y) \geq f(z)$ **then** $P = P \setminus \{z\} \cup \{y\}$ **end if**
 - 8: **end while**
-

Covantes Osuna and Sudholt (2019) based on the number of ones: $d(x, y) := ||x|_1 - |y|_1|$ where $|x|_1$ and $|y|_1$ denote the number of 1-bits in individual x and y , respectively.

2.1 Notation

Our notion of time is defined as the number of function evaluations before the $(\mu+1)$ EA achieves a stated goal such as finding a global optimum or finding both optima of TWO-MAX. Since the $(\mu+1)$ EA is initialised with μ individuals, and subsequently generates one offspring in each generation, the number of function evaluations is equal to μ plus the number of generations needed to achieve the set goal. The additional term of μ is only relevant for unreasonably large population sizes and is being tacitly ignored when it is absorbed in a runtime bound (such as $O(\mu n \log n)$) anyway.

We say that a function f is exponential if $f = 2^{\Omega(n^\varepsilon)}$ for a positive constant $\varepsilon > 0$. A function f is exponentially small if and only if $1/f$ is exponential. An event A occurs with overwhelming probability if $1 - \Pr(A)$ is exponentially small.

2.2 Drift Theorems

Our analysis will make heavy use of a technique called *drift analysis*. In a nutshell, the progress of the algorithm is measured by a potential function such as the Hamming distance to an optimum where a potential of 0 indicates that an optimum has been found. The *drift* is then defined as the expected change of this potential in one generation.

The following *multiplicative drift theorem* gives an upper bound on the expected time until the potential reaches 0 and an optimum has been found. It requires that the drift is at least proportional to its current state. It also gives a tail bound showing that the probability of exceeding this time is very small.

Theorem 1 (Multiplicative drift theorem with tail bounds, adapted from Doerr and Goldberg, 2013). *Let $\{X_t\}_{t \geq 0}$ be a sequence of random variables taking values in some set S . Let $g : S \rightarrow \{0\} \cup \mathbb{R}_{\geq 1}$ and assume that $g_{\max} := \max\{g(x) \mid x \in S\}$ exists. Let $T := \min\{t \geq 0 : g(X_t) = 0\}$. If there exists $\delta > 0$ such that*

$$\mathbb{E}[g(X_{t+1}) \mid g(X_t)] \leq (1 - \delta)g(X_t)$$

then $\mathbb{E}[T] \leq (1 + \ln g_{\max})/\delta$ and for every $c > 0$, $\Pr(T > (\ln g_{\max} + c)/\delta) \leq e^{-c}$.

3 Runtime Guarantees for Restricted Tournament Selection

We first provide runtime guarantees for restricted tournament selection, showing under which conditions and parameter settings RTS is efficient and inefficient, respectively.

E. Covantes Osuna and D. Sudholt

3.1 Large Window Sizes Are Effective

We start off by proving that the time bound of $O(\mu n \log n)$, used as measure of success in all previous runtime analyses of diversity mechanisms (cf. Table 1), applies to many variants of the $(\mu+1)$ EA. It assumes that the $(\mu+1)$ EA never decreases the best fitness on a considered branch of TWOMAX; we will show in the proof of Theorem 3 that this assumption is met with high probability in the context of RTS with large window sizes. We also use this time bound to explain why $O(\mu n \log n)$ constitutes a natural benchmark in the context of the $(\mu+1)$ EA enhanced with diversity mechanisms and to choose a sensible stopping criterion for our experimental analysis.

Lemma 2. *Consider one branch of TWOMAX and a $(\mu+1)$ EA with a replacement selection where the best fitness of all individuals on this branch never decreases and an offspring improving the current best fitness of its branch is always accepted³. If the $(\mu+1)$ EA is initialised with at least one individual on the branch then the optimum of the branch is found within time $\mu + 2e\mu n \ln n$ with probability $1 - 1/n$ and in expectation.*

Proof. We apply the multiplicative drift theorem with tail bounds (Doerr and Goldberg, 2010) (see Theorem 1) to random variables X_t that describe the Hamming distance of the closest individual to the targeted optimum. Note that $X_0 \leq n/2$ as we start with an individual on the considered branch and the optimum has been found once $X_t = 0$.

The probability of selecting an individual with Hamming distance X_t is at least $1/\mu$. In order to create a better individual, it is sufficient that one of the X_t differing bits is flipped and the other bits remain unchanged. Owing to our assumptions, such a better offspring will always survive. Each bit has a probability of being mutated of $1/n$ and the remaining bits remain unchanged with probability $(1 - 1/n)^{n-1} \geq 1/e$. Hence, the probability of creating an individual with a smaller Hamming distance is bounded as follows:

$$\Pr(X_{t+1} < X_t \mid X_t) \geq \frac{1}{\mu} \cdot \frac{X_t}{n} \cdot \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{X_t}{\mu en}.$$

This implies

$$\mathbb{E}[X_{t+1} \mid X_t] \leq \left(1 - \frac{1}{e\mu n}\right) X_t.$$

Applying Theorem 1 with $\delta = \frac{1}{e\mu n}$, $g_{\max} = n$, and starting with an individual with Hamming distance $< n/2$ to the optimum, yields that the time till the optimum is found is at most $e\mu n \cdot (\ln(n/2) + \ln n) \leq 2e\mu n \ln n$ with probability at most $1/n$ and in expectation. Adding a term of μ for the initial population completes the proof. \square

Now we state the main result of this section, a positive result for RTS when the window size w is large. The following analysis shows that, if w is chosen very large (even larger than the population size μ), the $(\mu+1)$ EA with RTS behaves almost like the $(\mu+1)$ EA with deterministic crowding.

Theorem 3. *If $\mu = o(\sqrt{n}/\log n)$ and $w \geq 2.5\mu \ln n$ then the $(\mu+1)$ EA with restricted tournament selection using genotypic or phenotypic distance finds both optima on TWOMAX in time $O(\mu n \log n)$ with probability at least $1 - 2^{-\mu'+3}$, where $\mu' := \min(\mu, \log n)$.*

³Compared to Lemma 3.3 in Covantes Osuna and Sudholt (2018b), the statement about offspring was added for clarification; it was implicitly assumed and used in Covantes Osuna and Sudholt (2018b). It is necessary as a branch might take longer to reach the optimum if better offspring are rejected; we will show later on, in Section 4.3, that this is the case for small window sizes w . In Theorem 3.1 of Covantes Osuna and Sudholt (2018b), the assumption was implicitly proven to hold for restricted tournament selection with large window sizes ($w \geq 2.5\mu \ln n$).

Note that, for small population sizes $\mu \leq \log n$, the probability $1 - 2^{-\mu'+3}$ is close to the success probability $1 - 2^{-\mu+1}$ for deterministic crowding (see Table 1 and Friedrich et al., 2009), apart from a constant factor in front of the $2^{-\mu'}$ term. For both, the success rate converges to 1 very quickly for increasing population sizes. For larger population sizes, $\mu > \log n$, the probability bound for restricted tournament selection is capped at $1 - 2^{-\log n+3} = 1 - 8/n$ as there is always a small probability of an unexpected takeover occurring.

In order to prove Theorem 3, we first analyse the probability of initialising a population such that there are individuals on each branch with a safety gap of σ to the border between branches. This safety gap will be used to exclude the possibility that the best individual on one branch creates offspring on the opposite branch.

Lemma 4. Consider the population of the $(\mu+1)$ EA on TWOMAX and for some μ and σ . The probability of having at least one initial search point with at most $n/2 - \sigma$ ones and one search point with at least $n/2 + \sigma$ ones is at least

$$1 - 2 \left(\frac{1 + 2\sigma \cdot \sqrt{2/n}}{2} \right)^\mu \geq 1 - 2^{-\mu+1} (1 + o(1))$$

where the inequality holds if $\sigma\mu = o(\sqrt{n})$.

Proof. By Doerr (2020, Equation 1.4.17), the following inequalities hold for any binomial coefficient:

$$\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor} \leq 2^n \cdot \sqrt{2/n}.$$

Hence, for a random variable with binomial distribution $\text{Bin}(n, 1/2)$, for all $z \in [0, n]$ we have

$$\Pr(X = z) \leq \Pr(X = \lfloor n/2 \rfloor) \leq 2^{-n} \cdot \binom{n}{\lfloor n/2 \rfloor} \leq \sqrt{2/n}.$$

So the probability that an individual x is initialised inside the safety gap is at most

$$p_\sigma := \Pr(n/2 - \sigma < |x|_1 < n/2 + \sigma) \leq 2\sigma \cdot \sqrt{2/n}.$$

Now let us define the probability that an individual x is initialised on the outer regions with $|x|_1 \leq n/2 - \sigma$ ones (0^n branch) or $|x|_1 \geq n/2 + \sigma$ ones (1^n branch) as p_0 and p_1 , respectively. Note that both p_0 and p_1 are symmetric, and $p_0 + p_1 := 1 - p_\sigma$, and by rewriting we obtain $p_0 := \frac{1-p_\sigma}{2}$ (the same for p_1) with its complement being $1 - \frac{1-p_\sigma}{2} = \frac{1+p_\sigma}{2}$.

So the probability of having no individual with at most $n/2 - \sigma$ ones is $(1 - p_1)^\mu = (\frac{1+p_\sigma}{2})^\mu$, and the same holds for having no individual with at least $n/2 + \sigma$ ones. Hence the probability of being initialised as stated in the statement of the lemma is at least

$$1 - 2 \left(\frac{1 + p_\sigma}{2} \right)^\mu = 1 - 2^{-\mu+1} \cdot (1 + p_\sigma)^\mu.$$

Plugging in p_σ and using the inequality $1 + x \leq e^x$ as well as $\sigma\mu = o(\sqrt{n})$ we simplify the last term as

$$(1 + p_\sigma)^\mu \leq e^{2\sigma\mu\sqrt{2/n}} = \frac{1}{e^{-o(1)}} \leq \frac{1}{1 - o(1)} = 1 + o(1),$$

and by plugging all together we have $1 - 2^{-\mu+1}(1 + o(1))$. □

E. Covantes Osuna and D. Sudholt

Using Lemmas 2 and 4, we can now prove Theorem 3.

Proof of Theorem 3. We first apply Lemma 4 with $\sigma := \log n$, noting that the assumption $\sigma\mu = o(\sqrt{n})$ holds true since we assume $\mu = o(\sqrt{n}/\log n)$. According to Lemma 4, with probability $1 - 2^{-\mu+1}(1 + o(1))$ the initial population contains at least one search point with at most $n/2 - \log n$ ones and at least one search point with at least $n/2 + \log n$ ones. We assume in the following that this has happened. Using $k! \geq (k/e)^k = k^{\Omega(k)}$, the probability of mutation flipping at least $\log n$ bits is at most $1/(\log n)! = (\log n)^{-\Omega(\log n)} = n^{-\Omega(\log \log n)}$. Taking the union bound over $O(\mu n \log n)$ steps still gives a superpolynomially small error probability. In the following, we work under the assumption that mutation never flips more than $\log n$ bits.

We call two search points *close* if their genotypic distance is at most $\log n$. Owing to our assumption on mutations, every newly created offspring is close to its parent. Note that on TWOMAX the phenotypic distance of any two search points is bounded from above by the genotypic distance, hence close search points also have a phenotypic distance of at most $\log n$. Note that, whenever the tournament contains a search point that is close to the new offspring, either the offspring or a close search point will be removed. If this always happens, the best individual on any branch cannot be eliminated by an offspring on the opposite branch; recall that initially, the best search points on the two branches have phenotypic distance at least $2 \log n$, and this phenotypic distance increases if the best fitness on any branch improves. When genotypic distances are being used, the genotypic distance is always at least $2 \log n$.

Since each offspring has at least one close search point (its parent), the probability that the tournament does not contain any close search point is at most $(1 - 1/\mu)^w \leq e^{-w/\mu}$. Using $w \geq 2.5\mu \ln n$, this is at most $e^{-2.5 \ln n} = 1/n^{2.5}$. So long as the best individual on any branch does not get replaced by any individuals on the opposite branch, the conditions of Lemma 2 are met. In particular, the assumption made in Lemma 2, that offspring improving the current best fitness of any branch are always accepted, holds true. Applying Lemma 2 to both branches, by the union bound the probability of both optima being found in time $2e\mu n \ln n$ is at least $1 - 2/n$. The probability that in this time a tournament occurs that does not involve a close search point is $O(\mu n \log n) \cdot 1/n^{2.5} = o(1/n)$ as $\mu = o(\sqrt{n}/\log n)$.

All failure probabilities sum up to (assuming n large enough)

$$\frac{2}{n} + o\left(\frac{1}{n}\right) + 2^{-\mu+1}(1 + o(1)) + \frac{O(\mu n \log n)}{n^{-O(\log \log n)}} \leq \frac{4}{n} + 2^{-\mu+2} \leq 2^{-\mu'+3}$$

where the last inequality follows as $2^{-\mu} \leq 2^{-\mu'}$ and $1/n \leq 2^{-\mu'}$. \square

In Theorem 3 we chose w so large that every tournament included the offspring's parent with high probability. Then the $(\mu+1)$ EA behaves like the $(\mu+1)$ EA with deterministic crowding (Friedrich et al., 2009), leading to similar success probabilities (see Table 1).

A success probability around $1 - 2^{-\mu+1}$ is best possible for many diversity mechanisms as with probability $2^{-\mu+1}$ the whole population is initialised on one branch only (for odd n), and then it is likely that only one optimum is reached. Methods like fitness sharing and clearing obtain success probabilities of 1 by more aggressive methods that can force individuals to travel from one branch to the other by accepting worse search points along the way. The performance of restricted tournament selection (and that of deterministic crowding) is hence best possible amongst all mechanisms that do not allow worse search points to enter the population.

The condition $w \geq 2.5\mu \ln n$ is chosen to ensure that the conditions of Lemma 2 are met with high probability in one generation, and throughout $O(\mu n \log n)$ generations. This condition is quite crude as it also means that in any fixed generation, the tournament will contain all individuals from the population with high probability. We believe that the algorithm also maintains niches on the two different branches with a much smaller choice of w as it is sufficient to have any close individual from the offspring's branch in the population to make sure that the offspring only competes against individuals from the same branch. We will investigate this experimentally in Section 5.

3.2 Small Window Sizes Can Fail

We now turn our attention to small w . If the w is small in comparison to μ , the possibility emerges that the tournament only contains individuals that are far from the offspring. In that case even the closest individual in the tournament will be dissimilar to the offspring, resulting in a competition between individuals from different "niches" (i. e., sets of similar individuals). The following theorem and its proof show that this may result in one branch taking over the other branch, even when the branch to get extinct is very close to a global optimum. The resulting expected runtime is exponential.

Theorem 5. *The probability that the $(\mu+1)$ EA with restricted tournament selection with $w \geq 3$ and either genotypic or phenotypic distances finds both optima on TWOMAX in time n^{n-1} is at most $1 - \exp\left(-\frac{2.28\mu^{w-1}}{n-1}\right) + O(1/n) = O(\mu^{w-1}/n)$. If $\mu \leq n^{1/(w-1)}$ then the expected time for finding both optima is $\Omega(n^n)$.*

For $w = 2$ the probability is at most $1 - \exp\left(-\frac{2\mu H_\mu}{n}\right) + O(1/n) = O((\mu \log \mu)/n)$, where H_μ refers to the μ -th harmonic number.

Theorem 5 shows an improvement over the preliminary version of this paper (Covantes Osuna and Sudholt, 2018b), which gave a weaker probability bound of $O(\mu^w/n)$.

Note that the probability of finding both optima in n^{n-1} generations is $o(1)$ if, for instance, $w = O(1)$ and μ grows slower than the polynomial $n^{1/(w-1)}$. It also holds if $w \leq c(\ln n)/\ln \ln n$ for some constant $0 < c < 1$ and $\mu = O(\log n)$ as then $n^{1/(w-1)} = e^{(\ln n)/(w-1)} \geq e^{(\ln \ln n)/c} = (\ln n)^{1/c} = \omega(\log n)$, which shows $\mu^{w-1}/n = o(1)$. The probability bound of $O(\mu^{w-1}/n)$ becomes trivial if $w \geq \log_\mu(n) + 1 = \log(n)/\log(w) + 1$ as then $\mu^{w-1}/n \geq 1$.

Proof of Theorem 5. We assume $\mu \leq n$ as otherwise all claimed probability bounds are larger than 1 for large enough n . The analysis follows the proof of Theorem 1 in Friedrich et al. (2009). We assume that the initial population contains at most one global optimum as the probability of both optima being found during initialisation is at most $\mu \cdot 2^{-n} \leq n \cdot 2^{-n}$, which can be easily subsumed in the terms of $O(1/n)$ in the claimed probability bounds.

We consider the first point of time at which the first optimum is being found. Without loss of generality, let us assume that this is 0^n . Then we show that with high probability copies of 0^n take over the population before the other optimum 1^n is found. The following arguments work for all populations that contain 0^n but not 1^n ; this includes the most promising population where all $\mu - 1$ remaining individuals are Hamming neighbours of 1^n .

Let i be the number of copies of the 0^n individuals in the population, then a good event G_i (good in a sense of leading towards extinction as we are aiming at a negative result) is to increase this number from i to $i + 1$. For this it is just necessary to create copies of one of the i individuals. For $n \geq 2$ we have $\Pr(G_i) \geq \frac{i}{\mu} \cdot \left(1 - \frac{1}{n}\right)^n \cdot \left(\frac{\mu-i}{\mu}\right)^w$

E. Covantes Osuna and D. Sudholt

since it suffices to select one out of i individuals and to create a copy of the selected individual, and to select w times individuals from the remaining $\mu - i$ individuals. On the other hand, a bad event B_i is to create an 1^n individual in one generation. This probability is clearly bounded by

$$\Pr(B_i) \leq \frac{\mu - i}{\mu} \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} + n^{-n}$$

as the probability to mutate 0^n into 1^n is n^{-n} and for the remaining $\mu - i$ individuals the chance of creating 1^n is at most $1/n \cdot (1 - 1/n)^{n-1}$. We deal with the term n^{-n} separately: the probability of such a jump happening in n^{n-1} steps is still at most $1/n$, which can be subsumed in the $O(1/n)$ term from the claimed failure probability. Hence we ignore this term in the following. Note that the quotient of both probability bounds is

$$\frac{\Pr(B_i)}{\Pr(G_i)} \leq \frac{\frac{\mu - i}{\mu} \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1}}{\frac{i}{\mu} \cdot \left(1 - \frac{1}{n}\right)^n \cdot \left(\frac{\mu - i}{\mu}\right)^w} = \frac{1}{n - 1} \cdot \frac{\mu^w}{i(\mu - i)^{w-1}}.$$

Together, the probability that the good event G_i happens before the bad event B_i is

$$\Pr(G_i \mid G_i \cup B_i) = \frac{\Pr(G_i)}{\Pr(G_i \cup B_i)} \geq 1 - \frac{\Pr(B_i)}{\Pr(G_i) + \Pr(B_i)} \geq \exp(-\Pr(B_i)/\Pr(G_i))$$

where the last step follows from the well-known inequality $1 + \frac{x}{1-x} \geq e^x$ for all $x < 1$ applied to $x := -\Pr(B_i)/\Pr(G_i)$.

The probability that the copies of 0^n take over the population before 1^n is found is therefore at least

$$\begin{aligned} \prod_{i=1}^{\mu-1} \Pr(G_i \mid G_i \cup B_i) &\geq \prod_{i=1}^{\mu-1} \exp\left(-\frac{\Pr(B_i)}{\Pr(G_i)}\right) \\ &= \exp\left(-\sum_{i=1}^{\mu-1} \frac{\Pr(B_i)}{\Pr(G_i)}\right) \geq \exp\left(-\frac{\mu^w}{n-1} \sum_{i=1}^{\mu-1} \frac{1}{i(\mu-i)^{w-1}}\right). \end{aligned} \quad (1)$$

For $w = 2$ the last sum simplifies to

$$\sum_{i=1}^{\mu-1} \frac{1}{i(\mu-i)} = \sum_{i=1}^{\mu-1} \left(\frac{1}{i\mu} + \frac{1}{\mu(\mu-i)}\right) = \frac{2H_{\mu-1}}{\mu}$$

and the probability that takeover happens is at least

$$\exp\left(-\frac{\mu^w}{n-1} \cdot \frac{2H_{\mu-1}}{\mu}\right) \geq \exp\left(-\frac{2\mu H_{\mu-1}}{n}\right)$$

where the inequality holds since $w = 2$ and $(\mu - 1)/(n - 1) \leq \mu/n$, which in turn is implied by $\mu \leq n$. Along with the error term of $O(1/n)$, this yields the claimed probability bound for $w = 2$.

For $w \geq 3$ we note that the summands in (1) are non-increasing with w . So the worst case is having the smallest possible value, $w = 3$. Note that

$$\frac{1}{i(\mu-i)^2} = \frac{\mu-i}{i\mu(\mu-i)^2} + \frac{i}{i\mu(\mu-i)^2} = \frac{1}{i\mu(\mu-i)} + \frac{1}{\mu(\mu-i)^2} = \frac{1}{\mu} \left(\frac{1}{i(\mu-i)} + \frac{1}{(\mu-i)^2}\right).$$

Thus

$$\begin{aligned} \sum_{i=1}^{\mu-1} \frac{1}{i(\mu-i)^2} &= \frac{1}{\mu} \left(\sum_{i=1}^{\mu-1} \frac{1}{i(\mu-i)} + \sum_{i=1}^{\mu-1} \frac{1}{(\mu-i)^2} \right) \\ &= \frac{1}{\mu} \left(\frac{2H_{\mu-1}}{\mu} + \sum_{i=1}^{\mu-1} \frac{1}{(\mu-i)^2} \right) \\ &= \frac{2H_{\mu-1}}{\mu^2} + \frac{1}{\mu} \sum_{i=1}^{\mu-1} \frac{1}{i^2}. \end{aligned}$$

Since $\sum_{i=1}^{\mu-1} 1/i^2 \leq \sum_{i=1}^{\infty} 1/i^2 = \pi^2/6 \approx 1.645$, the above approaches $1.645/\mu$ as μ grows. The function is $2/\mu$ for $\mu = 2$, $2.25/\mu$ for $\mu = 3$, and at most $2.28/\mu$ for $\mu \geq 4$, with the constant factor decreasing with increasing μ for $\mu \in [4, \infty)$. Hence the function is bounded by $2.28/\mu$ for all $\mu \in \mathbb{N}$.

Together we have

$$\prod_{i=1}^{\mu} \Pr(G_i \mid G_i \cup B_i) \geq \exp\left(-\frac{\mu^w}{n-1} \cdot \frac{2.28}{\mu}\right) = \exp\left(-\frac{2.28\mu^{w-1}}{n-1}\right) \geq 1 - O(\mu^{w-1}/n).$$

Once the population consists only of copies of 0^n , a mutation has to flip all n bits to find the 1^n optimum. This event has probability n^{-n} and, by the union bound, the probability of this happening in a phase consisting of n^{n-1} generations is at most $\frac{1}{n} = O(\mu^{w-1}/n)$. The sum of all failure probabilities is $O(\mu^{w-1}/n)$, which proves the first claim. For the second claim, observe that the conditional expected runtime is n^n once the population has collapsed to copies of 0^n individuals. Using $\mu \leq n^{1/(w-1)}$ this situation occurs with probability at least $\exp\left(-\frac{2.28n}{n-1}\right) - O(1/n) = \Omega(1)$. Hence the unconditional expected runtime is $\Omega(n^n)$. \square

4 Runtime Guarantees for RTS without Replacement

In this section we now discuss what happens if RTS is modified to select w individuals *without replacement*, that is, w different individuals are being selected. We call this a modification as we believe the original RTS selects with replacement (this is not mentioned explicitly, but it follows from the mathematical formulae in Harik (1995)). Selecting without replacement makes as much sense as selecting with replacement, and it is very plausible that many practical implementations have used one or the other variant. Selecting without replacement leads to a more diverse tournament and we expect a stronger effect, compared to selecting with replacement, when the same window size is used. Note that if $w = \mu$ then the whole population is selected for the tournament and thus a closest individual in the population is selected to compete against the offspring. We define the algorithm for $w > \mu$ as well, for consistency with experiments for RTS with replacement (that use values for w larger than μ), even though there is no difference to $w = \mu$. If $w > \mu$ we select the whole population and $w - \mu$ copies of arbitrary individuals; the effect is the same as for $\mu = w$.

4.1 Large Window Sizes Still Work

The positive result mentioned in Section 3.1 still applies for this variant of RTS. The main difference to Theorem 3 is that, assuming the algorithm never flips at least $\log n$

E. Covantes Osuna and D. Sudholt

bits, if $w \geq \mu$ the $(\mu+1)$ EA always selects at least one close search point for the tournament (as opposed to a probability of at least $1 - 1/n^{2.5}$). We thus obtain the same result as Theorem 3 for the weaker condition $w \geq \mu$. Note that this condition implies that the whole population is contained in the tournament, and so the offspring always competes against the closest individual from the population.

Theorem 6. *If $\mu = o(\sqrt{n}/\log n)$ and $w \geq \mu$ the $(\mu+1)$ EA with restricted tournament selection selecting w individuals without replacement, using genotypic or phenotypic distance, finds both optima on TWOMAX in time $O(\mu n \log n)$ with probability at least $1 - 2^{-\mu'+3}$, where $\mu' := \min(\mu, \log n)$.*

4.2 On Takeover with Small Window Sizes

Unlike our positive result that works for both selection policies, the negative result (Theorem 5) is no longer applicable for RTS without replacement. Recall that for the negative result we rely on an extreme case where the algorithm has found the first optimum 0^n and then this optimum takes over the whole population before the opposite optimum 1^n is found. Consider the situation where there is just one individual x left (a Hamming neighbour of 1^n) before copies of 0^n take over and the new offspring y is another copy of 0^n . Theorem 5 then relied on individual x being selected w times for the tournament in order to complete the takeover as then x is the closest individual to y . When using RTS without replacement, this event is impossible for $w \geq 2$ as the tournament is guaranteed to contain a copy of 0^n , which then competes against the new offspring y . Hence, in this case, the subpopulation on the branch towards 1^n will never become extinct and will eventually reach 1^n .

In fact, the following lemma shows that if the $(\mu+1)$ EA using RTS without replacement and phenotypic distances is able to maintain subpopulations on both branches until both subpopulations have evolved to a fitness larger than $2n/3$, extinction becomes impossible and it is certain that both optima will be found eventually. For genotypic distances a similar statement holds when all individuals have passed a (higher) fitness threshold of $3n/4$.

Lemma 7. *Consider the $(\mu+1)$ EA using restricted tournament selection without replacement and $w \geq 2$. If phenotypic distances are used, once a population is reached that has subpopulations on both branches whose best fitness is larger than $2n/3$, no subpopulation will become extinct. If genotypic distances are used, once a population is reached that has subpopulations on both branches whose worst fitness is larger than $3n/4$, no subpopulation will become extinct.*

Proof. A subpopulation can only become extinct if it only contains a single individual. We hereinafter call this individual x and observe that by assumption $f(x) > 2n/3$. Without loss of generality, we assume that x is on the 1-branch, thus $|x|_1 > 2n/3$. The following events are necessary for extinction: an offspring y of fitness $f(y) \geq f(x)$ is created on the opposite branch to x (i. e., $|y|_1 < n/3$), the single individual x is chosen for the tournament, and x is the closest search point to the offspring y . Only then will x be removed from the population.

We show that these conditions are impossible. The phenotypic distance between x and the offspring y is larger than $n/3$. Since $w \geq 2$, the tournament must also contain an individual x' from the 0-branch. If $f(x') < f(y)$ then x' is phenotypically closer to y than x . Otherwise, since $0 \leq |x'|_1 \leq |y|_1 < n/3$, the phenotypic distance between x' and the offspring y is less than $n/3$, hence x cannot be the closest search point to the offspring y . Consequently, x cannot be replaced by an offspring on the opposite branch.

With genotypic distances the same arguments apply. Assume that $|x|_1 > 3n/4$ and $f(y) \geq f(x)$, i. e., $|y|_1 < n/4$. Then the genotypic distance between every individual x' from the 0-branch and the offspring y is less than $n/2$ as by assumption $|x'|_1 < n/4$ and thus both y and x' have Hamming distance less than $n/4$ to 0^n , whereas the genotypic distance between x and the offspring y is larger than $n/2$. \square

In scenarios where the fitness thresholds from Lemma 7 are not reached, takeover of one branch may still happen. Imagine a population where an offspring y is created on one branch and the tournament contains an individual z from the opposite branch with $f(z) < f(y)$. If the tournament only contains other search points whose distance to y is larger than the distance between z and y , z will be removed from the population. If such steps happen repeatedly, the subpopulation on z 's branch may become extinct. Note, however, that if the size of said subpopulation is less than w , the tournament must contain individuals from y 's branch that have a large distance from y .

Figure 1 sketches a population described above. The population is divided into two species. On the 1-branch there is only one individual (orange point in Figure 1 or individual z) and on the 0-branch there are $\mu - 1$ individuals and the offspring (red point in Figure 1 or individual y), which has a better fitness than z . The parent individuals on the 0-branch contain a subset P' of individuals whose distance to y is larger than the distance between z and y . In this case, if the tournament consists of z and $w - 1$ individuals from P' then z is the closest individual to y and y replaces z .

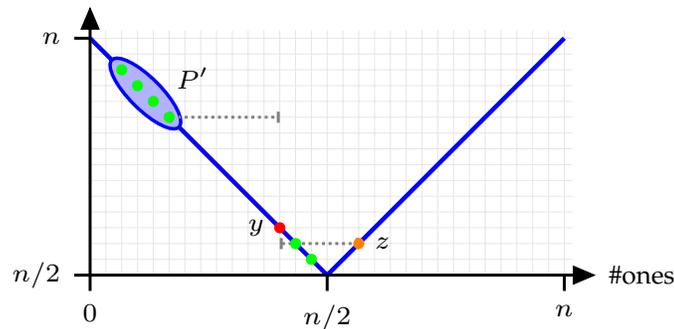


Figure 1: Sketch of $f = \text{TWOMAX}$ with a population where takeover may happen for the $(\mu+1)$ EA with RTS selecting without replacement, using phenotypic distances. The offspring (y , red) replaces the single individual (z , orange) if the tournament contains z and $w - 1$ individuals from P' as then z is the one closest individual from the tournament as indicated by the dotted lines.

Note that the scenario exemplified in Figure 1 is impossible under the conditions from Lemma 7 as there z would have a much higher fitness and the niche P' with the mentioned properties cannot exist. It is also highly unlikely for such a population to emerge in the $(\mu+1)$ EA without any diversity mechanisms as the subpopulations on both branches tend to have similar fitness values. When using RTS, however, it is possible for different niches like P' , and niches with significantly worse fitness, to emerge. We will further investigate the likelihood of takeover in the experimental analysis (Section 5).

E. Covantes Osuna and D. Sudholt

4.3 RTS with Small Window Sizes Slows Down Evolution

As mentioned in previous sections, RTS without replacement is proven to be effective for smaller values of w . But this comes at a price as the time to reach both optima can increase significantly. In a scenario where there is just one individual x on one branch, this “lone” individual will need to climb up its branch until it reaches its respective optimum. Assume we have a generation where x was selected as a parent and has produced a better offspring on the same branch. Further assume that x and its offspring have a worse fitness than all individuals on the opposite branch. Then the only way that the subpopulation of x can evolve is if x is selected in the tournament. In this case, x will compete against its (fitter) offspring and will be removed from the population. In other words, x has to be removed to allow its offspring to survive. This means that x will need to be selected twice in one generation: as parent and in the replacement selection. Also note that the new subpopulation will consist of just one individual, hence the algorithm may be in the same situation for a long period of time.

We make this precise in the following lemma which shows that the evolution on a branch with only a single individual proceeds as in a *lazy*⁴ (slowed-down) (1+1) EA. In addition to assuming a branch with a single individual x , we assume that all individuals on the opposite branch will have a fitness that is larger than the fitness of x by an amount that is at least logarithmic.

Definition 8. Define the p -lazy (1+1) EA as an algorithm that independently in each iteration idles with probability p and otherwise (that is, with probability $1 - p$) performs one step of the (1+1) EA.

The following lemma assumes without loss of generality that the lone individual resides on the 1-branch. A symmetric statement holds when the roles of the two branches are swapped.

Lemma 9. Consider the $(\mu+1)$ EA using restricted tournament selection without replacement, window size $w \leq \mu$ and genotypic or phenotypic distance. Suppose that the population P_t at time t contains a single individual x on the 1-branch with fitness at least $n/2 + \log n$ and that all other individuals in P_t are on the 0-branch and have a fitness of at least $f(x) + \log n$. Then, with probability $1 - n^{-\omega(1)}$, the subpopulation on the 1-branch in the $(\mu+1)$ EA will evolve as in one step of the (w/μ^2) -lazy (1+1) EA with current search point x on ONEMAX.

Proof. If any of the search points on the 0-branch is selected as parent, at least $\log n$ bits have to flip to create a search point on the 1-branch of fitness at least $f(x)$. This has probability $n^{-\omega(1)}$ and it is a necessary condition for the subpopulation on the 1-branch to change (as a worse individual on the 1-branch will be removed regardless of the outcome of the tournament).

If x is selected as parent, at least $\log n$ bits have to flip to create a search point that is on the 0-branch or at least as good as the worst individual on the 0-branch. This, again, has probability $n^{-\omega(1)}$. If this does not happen, the offspring y can only survive if the tournament contains x and $f(y) \geq f(x)$. In this case, y replaces x . The probability for the tournament containing x is w/μ (there are $\binom{\mu}{w}$ ways of choosing w individuals without replacement and $\binom{\mu-1}{w-1}$ ways of choosing x and $w - 1$ other individuals from the remaining $\mu - 1$ individuals; together, this yields a probability of $\binom{\mu-1}{w-1} / \binom{\mu}{w} = w/\mu$). Along with a probability of $1/\mu$ for selecting x as parent, the probability of y replacing x in case $f(y) \geq f(x)$ is w/μ^2 . \square

⁴In Markov chain theory, a Markov chain is called *lazy* if there is a fixed probability (e. g., $1/2$) of remaining in the same state (Levin et al., 2009).

Lemma 9 (along with the well-known fact that the (1+1) EA requires $\Theta(n \log n)$ time on ONEMAX) suggests that, under appropriate conditions, a lower bound of $\Omega((\mu^2/w) \cdot n \log n)$ applies for the $(\mu+1)$ EA with RTS with replacement on TWOMAX. However, to formally prove such a bound, we would need to show that the assumptions of the lemma have a good chance to be satisfied during an appropriate time period.

In preliminary experiments we observed that, for $w = 2$ and across a range of values for μ , a lone individual emerged in almost all runs. Furthermore, in those cases the individuals on the opposite branch evolved faster, and the time to find both optima was determined by the time the lone individual evolved its respective optimum. We do not have a formal proof that the larger subpopulation evolves faster, though. And it is not clear whether a lone individual typically develops for larger values of w . We therefore resort to experiments in Section 5 to investigate this matter further and to check for which values of w our conjectured lower bound of $\Omega((\mu^2/w) \cdot n \log n)$ is supported.

5 Experimental Analysis

We provide an experimental analysis as well in order to see how closely the theory matches the empirical performance for a reasonable problem size, and to investigate a wider range of parameters, where the theoretical results are not applicable. We are interested in the impact of the window size w and the selection policy (with or without replacement) on the success rate of the $(\mu+1)$ EA with RTS.

Another interesting question is to compare RTS with and without replacement in its resilience to takeover. We argued in Section 4 that the RTS without replacement is more resilient to takeover, compared to its variant with replacement, but at the expense of an increased runtime for finding both optima. We also argued that there are scenarios where takeover may happen when no replacement is used, so we would like to know for how long RTS is able to delay takeover and when it is more likely that takeover happens for both selection policies.

We consider exponentially increasing population sizes $\mu \in \{2, 4, 8, \dots, 1024\}$ for a problem size $n = 100$ and for 100 runs. Based on our theoretical analysis we define the following outcomes and stopping criterion for each run. *Success*, the population contains both 0^n and 1^n in the population. And *failure*, once the run has reached a maximum number of generations and the population does not contain both optima. This maximum is initially set to $10\mu n \ln n$ as motivated by Lemma 2, with a more generous leading constant that leaves plenty of time for most diversity mechanisms to find both optima in case no takeover happens. The same time bound was also used in Covantes Osuna and Sudholt (2018a), an empirical comparison of a range of diversity mechanisms. In Section 5.2 we will consider a larger time budget to reflect our conjecture that RTS without replacement requires more time by a factor of order μ^2/w , for appropriate values of w .

For both RTS variants, we tested window sizes $w \in \{1, 2, 4, 8, \dots, 1024\}$, however we only plot results up to $w = 128$ as the results for large w were very similar. In the particular case where $w > \mu$ and the tournament is selected without replacement, the algorithm uses the whole population in the tournament.

5.1 Experiments for a Time Budget of $10\mu n \ln n$

In Figures 2a and 2b we show the results for the $(\mu+1)$ EA with the original RTS, selecting w individuals with replacement. For small values of w and μ the algorithm is not able to maintain individuals on both branches of TWOMAX for a long period of time,

E. Covantes Osuna and D. Sudholt

as predicted by Theorem 5. It is only when the population size is set to $\mu = 1024$ and $w = 1$ that the algorithm is able to maintain individuals on both branches before the takeover happens. When setting, for example, $w \geq 8$ and $\mu \geq 32$, the algorithm was able to find both optima with both genotypic and phenotypic distances. It is possible to observe a trade-off between w and μ : a larger w allows for a smaller population size μ to be used. Such a trade-off was also indicated by the probability bound $O(\mu^{w-1}/n)$ from Theorem 5.

Our experiments show that RTS works well for much smaller window sizes than those required in Theorem 3. For instance, for $\mu = 32$ Theorem 3 results a window size of $w \geq 369$, whereas Figure 2d shows that $w \geq 8$ seems sufficient to be effective. Note that the method seems to behave fairly similarly with respect to both distance functions.

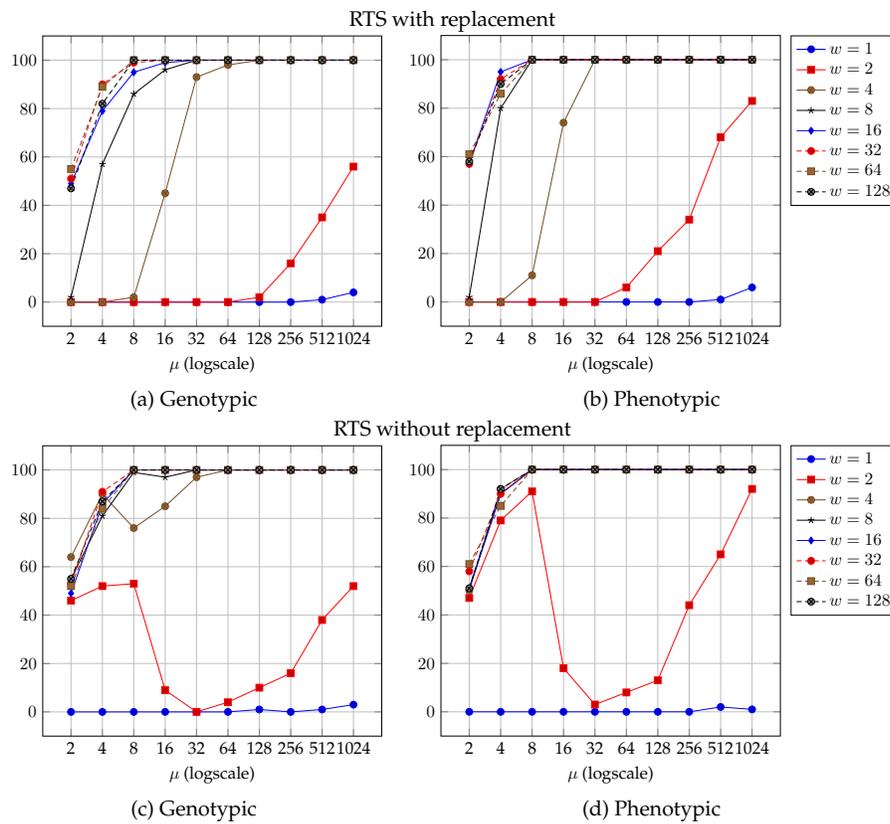


Figure 2: The number of successful runs measured among 100 runs at the time both optima were found on TWOMAX or $t = 10\mu n \ln n$ generations have been reached for $n = 100$ with the $(\mu+1)$ EA with restricted tournament selection with and without replacement for $\mu \in \{2, 4, 8, \dots, 1024\}$, $w \in \{1, 2, 4, 8, \dots, 128\}$, genotypic and phenotypic distance.

Now, Figures 2c and 2d illustrate the performance of RTS without replacement. While all curves for RTS with replacement and $w \geq 8$ in Figure 2c were monotonically increasing, the success rate for $w = 2$ without replacement in Figure 2d is clearly not monotonic. For population sizes $\mu \in \{2, 4, 8\}$ the success rate does increase, but once

the population size increases to $\mu = 16$ or $\mu = 32$, the success rate drops steeply. Finally, the success rate starts going up again when $\mu \geq 64$.

From our theoretical considerations and observing individual runs, the reason for this drop in the success rate seems to be due to the scenarios described in Section 4.3. During the runs, if there is a subpopulation of individuals with better fitness than the individuals on the opposite branch of TWOMAX, the better subpopulation starts taking over until there is just one individual x in the worse subpopulation. With $w = 2$ and using selection without replacement, in order for an improving mutation y of x to be accepted, the tournament must contain x , assuming all individuals in the better subpopulation have a higher fitness than y . Hence, for the worse subpopulation of one individual to progress, the single individual must be selected as parent and also be selected for the tournament as deleting x is the only way for its offspring to survive. Compared to a setting with large window sizes (Theorems 3 and 6), this indicates an additional factor of μ/w in the expected time for the worst subpopulation to reach its optimum (cf. Lemma 9). Hence the drop in success rates seems to be caused by the time budget not being large enough to allow a single individual on one branch to evolve into its respective optimum (we will investigate this further in the following by increasing the time budget). This behaviour only occurs for intermediate population sizes μ as for small μ the difference between factors of μ and μ^2/w is not significant enough (recall that we have chosen a generous leading constant of 10 in the time bound $10\mu n \ln n$). For large μ , the algorithm is efficient, albeit less efficient than for larger values of the window size w .

5.2 An Increased Time Budget of $100 \max(1, \mu/w)\mu n \ln n$

Since we hypothesise that RTS without replacement requires more time to achieve positive results (see the theoretical results from Section 4.3, particularly Lemma 9), we performed the same experiments but increasing the time bound by an additional factor of $\max(1, \mu/w)$, where the maximum is used to accommodate values of $w > \mu$ that show identical behaviour to $w = \mu$. We also increased the leading constant from 10 to 100 to ensure that success or failure does happen in the given time, resulting in an enhanced time budget of $100 \max(1, \mu/w)\mu n \ln n$. Our experiments will confirm that this budget was sufficient in all runs.

In the case of RTS with replacement with the new time bound (Figures 3a and 3b), there are no major changes compared to the previous results shown in Figures 2a and 2b. This suggests that the original time budget is sufficient for the algorithm to arrive at the defined outcomes (success or failure). The major difference is shown with respect to RTS without replacement with the new time budget (Figures 3c and 3d) compared to the same algorithm with the original time budget (Figures 2c and 2d). As can be observed the performance shown in Figures 3c and 3d show that RTS without replacement has a better performance for small values $w \geq 2$ and $\mu \geq 16$ when given a larger time budget.

These results support our hypothesis that RTS without replacement has a better success rate than its variant with replacement for much smaller w and μ , but that this comes at the expense of a higher runtime as explained in Section 4.3. In this case takeover is not possible for $w = 2$, but as can be seen from Figures 3c and 3d, evolving a single individual through the whole branch takes more time, which explains the sudden drop in the performance shown in Figures 2c and 2d. This aligns with our hypothesis that the number of evaluations increases by a factor of μ^2/w to find both optima on TWOMAX. Finally, aside from the extreme case where $w = 2$, the algorithm

E. Covantes Osuna and D. Sudholt

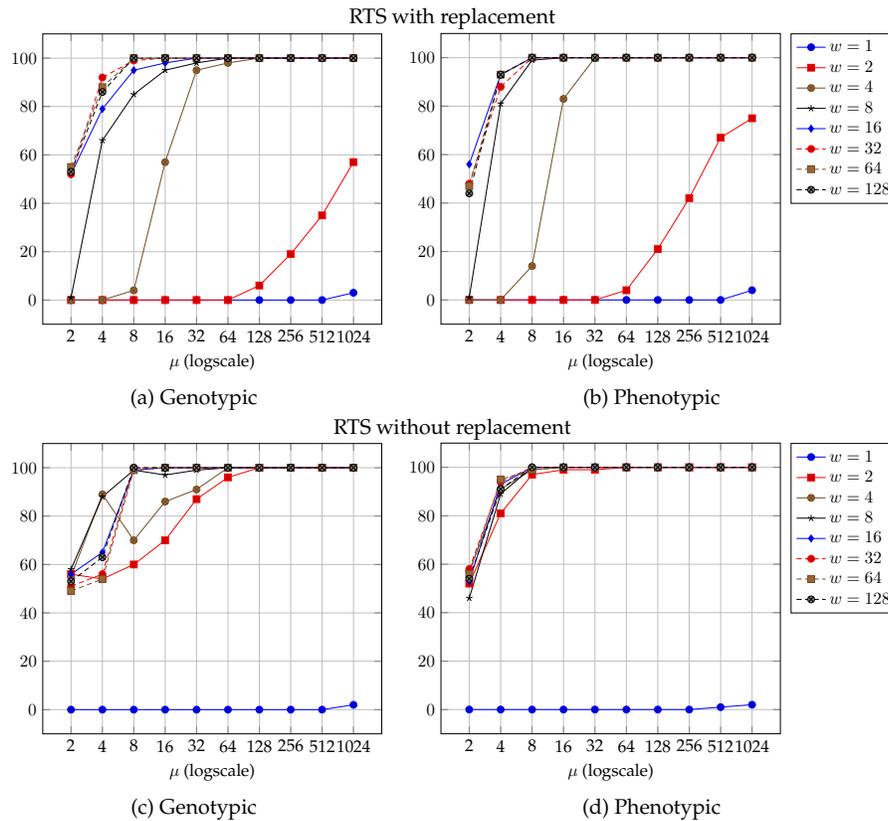


Figure 3: The number of successful runs measured among 100 runs at the time both optima were found on TWOMAX or $t = 100 \max(1, \mu/w) \mu n \ln n$ generations have been reached for $n = 100$ with the $(\mu+1)$ EA with restricted tournament selection with and without replacement and $\mu \in \{2, 4, 8, \dots, 1024\}$, $w \in \{1, 2, 4, 8, \dots, 128\}$, genotypic and phenotypic distance.

seems to behave similarly with both time budgets, achieving good results when $w \geq 8$ and $\mu \geq 8$ and achieving 100% success rate when $w \geq 8$ and $\mu \geq 32$. The choice $w = 4$ gives mixed results as for both time budgets the success rate for RTS without replacement and genotypic distance is not monotonic in μ .

5.2.1 Slow Down in the Performance of RTS Without Replacement on TWOMAX

Now we look more closely into the slow down in the performance of RTS without replacement due to the appearance of the lone individual. We performed new experiments where we removed the time budget of generations, and we defined as the only stopping criterion either finding both optima, or that the population consists of copies of one optimum and with $w - 1$ individuals with fitness $n - 1$.

The reason for this slightly different stopping criterion is due to the selection without replacement of RTS. When the population has collapsed into one branch of TWOMAX, and the population has reached the optimum, once there are $w - 1$ individuals in the population with fitness $n - 1$, we claim that the algorithm is no longer able to replace the $w - 1$ individuals with fitness $n - 1$. Imagine that another copy of the present

optimum is created. Since there are only $w - 1$ non-optimal individuals, any tournament of size w must contain an optimum that is identical to the offspring. Then the new offspring will automatically compete with a global optimum since it is the closest individual in the tournament. In this sense, unless the opposite optimum is created by mutation, only replacements amongst optimal individuals are possible and the $w - 1$ individuals with fitness $n - 1$ will remain untouched. The algorithm will idle forever since it is not possible for all μ individuals to reach a fitness value of n . So once the algorithm reaches this “stagnation” scenario the run is stopped since there is no way to introduce new changes in the population.

In Figure 4 we show how much time is needed by reporting the average number of generations achieved using the stopping criterion defined previously for the case of RTS without replacement with genotypic distance. An extra feature in this experimental setting is that we have introduced an additional curve for $w = \mu$ in order to observe whether the average runtime grows with μ^2/w . Together with Figure 4, we provide in Table 2 the mean and standard deviation of generations required for the same experimental setting as Figure 4, and the number of times the lone individual scenario came up among the 100 runs of the RTS without replacement.

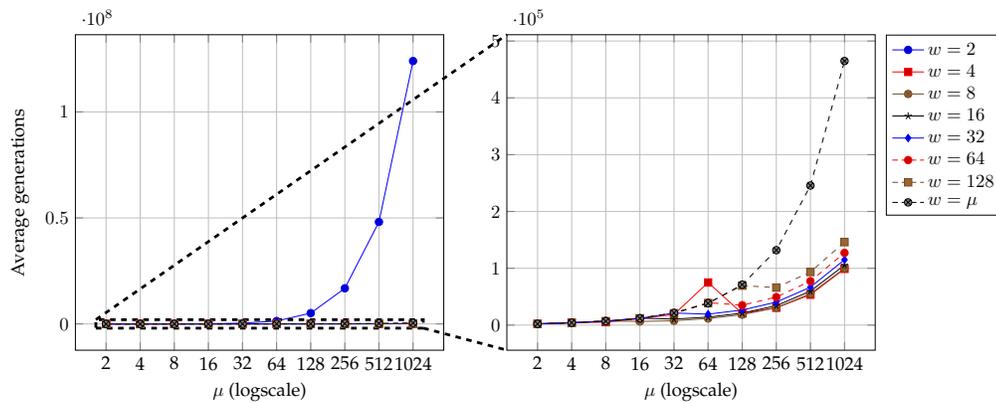


Figure 4: Average number of generations until either both optima of TWOMAX are found or the population consists of copies of the optimum and with $w - 1$ individuals with fitness $n - 1$ among 100 runs for $n = 100$ with the $(\mu+1)$ EA with restricted tournament selection without replacement for $\mu \in \{2, 4, 8, \dots, 1024\}$, $w \in \{2, 4, 8, \dots, 128, \mu\}$ and genotypic distance. The right-hand side shows a close-up view of the plot on the left-hand side (indicated by the dashed lines), without the curve for $w = 2$.

First of all, from the raw data we observed that the increased time budget of $t := 100 \max(1, \mu/w) \mu n \ln n$ was never reached, hence it seems to be large enough, and effectively the RTS without replacement needs more time to find both optima on TWOMAX. From Figure 4, when $w = 2$, we can observe that the average runtime immediately increases as soon as the population size increases. Note that in Table 2 for $w = 2$ lone individuals appear frequently. This supports our hypothesis from Section 4.3 that due to the lone individual appearance during the run, there is an increase in the mean of generations needed to find their respective optimum. The mean increases when increasing the population size μ and the standard deviation increases in line with the mean. In most runs, the algorithm is idling as the lone individual needs to be selected for both selection for reproduction and selection for the tournament.

E. Covantes Osuna and D. Sudholt

Table 2: Mean and std of generations required to either find both optima of TWOMAX or the population consists of copies of the optimum and with $w - 1$ individuals with fitness $n - 1$, and the number of lone individual appearances among 100 runs for $n = 100$ with the $(\mu+1)$ EA with restricted tournament selection without replacement for $\mu \in \{2, 4, 8, \dots, 1024\}$, $w \in \{2, 4, 8, \dots, 128, \mu\}$ and genotypic distance.

μ		w							μ
		2	4	8	16	32	64	128	
2	mean	2.22E+03	2.20E+03	2.07E+03	2.16E+03	2.22E+03	2.23E+03	2.20E+03	2.18E+03
	std	6.31E+02	6.18E+02	5.71E+02	6.36E+02	6.45E+02	6.40E+02	5.36E+02	5.90E+02
	lone	55	52	55	52	52	50	47	60
4	mean	6.84E+03	4.07E+03	4.12E+03	4.15E+03	4.20E+03	4.21E+03	4.06E+03	4.28E+03
	std	1.96E+03	9.05E+02	9.12E+02	1.12E+03	1.03E+03	8.09E+02	9.08E+02	9.81E+02
	lone	88	64	68	60	57	71	65	66
8	mean	2.35E+04	5.28E+03	7.15E+03	7.00E+03	7.08E+03	7.17E+03	6.95E+03	6.90E+03
	std	6.23E+03	3.11E+03	1.39E+03	1.65E+03	1.49E+03	1.35E+03	1.32E+03	1.65E+03
	lone	97	44	10	6	6	8	14	7
16	mean	9.02E+04	1.13E+04	6.50E+03	1.22E+04	1.19E+04	1.20E+04	1.20E+04	1.19E+04
	std	2.58E+04	1.09E+04	7.29E+03	1.57E+03	1.65E+03	1.61E+03	1.84E+03	1.69E+03
	lone	100	21	4	0	1	0	0	0
32	mean	3.67E+05	1.89E+04	7.62E+03	1.06E+04	2.13E+04	2.13E+04	2.08E+04	2.14E+04
	std	1.10E+05	3.39E+04	1.12E+03	1.53E+03	2.39E+03	2.38E+03	2.27E+03	2.36E+03
	lone	100	7	1	0	0	0	0	0
64	mean	1.37E+06	7.50E+04	1.16E+04	1.41E+04	1.93E+04	3.95E+04	3.88E+04	3.84E+04
	std	4.14E+05	5.98E+05	2.39E+03	1.71E+03	2.51E+03	4.03E+03	4.43E+03	3.97E+03
	lone	100	1	0	0	0	0	0	0
128	mean	5.10E+06	2.05E+04	1.83E+04	2.11E+04	2.66E+04	3.51E+04	6.95E+04	7.12E+04
	std	1.95E+06	5.42E+03	1.44E+03	1.94E+03	2.44E+03	3.71E+03	6.70E+03	6.01E+03
	lone	96	0	0	0	0	0	0	0
256	mean	1.68E+07	3.05E+04	3.07E+04	3.42E+04	4.05E+04	4.93E+04	6.60E+04	1.32E+05
	std	8.79E+06	4.99E+03	2.05E+03	2.41E+03	3.48E+03	3.59E+03	6.58E+03	1.01E+04
	lone	88	0	0	0	0	0	0	0
512	mean	4.81E+07	5.35E+04	5.54E+04	5.97E+04	6.68E+04	7.74E+04	9.34E+04	2.46E+05
	std	3.89E+07	5.26E+03	3.15E+03	3.98E+03	4.82E+03	5.51E+03	7.22E+03	1.80E+04
	lone	70	0	0	0	0	0	0	0
1024	mean	1.24E+08	9.90E+04	1.00E+05	1.06E+05	1.15E+05	1.27E+05	1.46E+05	4.65E+05
	std	1.42E+08	8.76E+03	5.56E+03	5.64E+03	6.86E+03	8.32E+03	1.00E+04	3.64E+04
	lone	50	0	0	0	0	0	0	0

In the case of $4 \leq w \leq \mu$ we can observe an improvement in the performance of the algorithm. The chances of spending time idling for a significant replacement decrease since we are allowing more individuals to participate in the tournament. From both, Figure 4 and Table 2, for $w = 4$ we still can see some of this idling behaviour with fewer appearances of the lone individual as the population size increases. However, the means and standard deviations are generally smaller than for $w = 2$. For the case of $8 \leq w \leq 128$ the algorithm seems to be much faster than predicted by our conjectured bound $\Omega((\mu^2/w) \cdot n \log n)$. A plausible explanation drawn from Table 2 is that for $w \geq 8$ the chances of lone individuals emerging are much lower than for $w \in \{2, 4\}$.

Finally, for the case of $w = \mu$ there is a slow down towards the optimum because the whole population is part of the tournament, and only replacements can be done with individuals close to each other, which translates into small jumps and more time needed to move towards the optimum. In this case the effect of RTS vanishes and the

size of the population is the only factor in the growth of the time, something that we were expecting from the term μ^2/w in our conjectured bound of $\Omega((\mu^2/w) \cdot n \log n)$.

It seems that our conjectured bound of $\Omega((\mu^2/w) \cdot n \log n)$ only applies to very small values of w due to the appearance of the lone individual and the necessity of specific selections in one step. For the case of large values of $w = \mu$, the RTS effect vanishes and we end up with an algorithm similar to deterministic crowding in which the only main factor on the runtime is the population size. For intermediate values of w , it seems that a more relaxed time bound can be applied since unwanted replacements are more difficult (more individuals participate in the tournament) and specific selections for replacement are not needed (large jumps in the same branch are possible) but this may need more detailed conditions or arguments related to the population dynamics and distribution on the population.

5.3 Progress on the Inferior Niche

The main goal of any niching mechanism is to avoid (or at least delay) that the fittest individuals take over the less fit individuals, which ultimately leads to a loss of diversity. From the previous theoretical and empirical analysis we can observe that RTS is able to delay takeover depending on the selection scheme, and the parameters w and μ . We know that RTS with replacement, for small values of w and μ , is not able to maintain individuals on both branches of TWOMAX for a long period of time. For the case of RTS without replacement we know that takeover it is more difficult to happen, the algorithm is more resilient to takeover, but as we mentioned in Section 4.2, takeover is still possible.

So, when exactly is takeover more likely to happen? When exactly during the process does a subpopulation become extinct? When and under what situations does extinction happen? These are the kind of questions we aim to answer in this section. In order to observe when takeover is more likely to happen with both variants of RTS, we have designed the following experimental analysis. We recorded the fitness of the best individuals reached on both branches of TWOMAX, and then take the minimum fitness value of those two individuals. This yields the maximum fitness reached by the subpopulation that becomes extinct (if applicable), or the optimal fitness if both optima were reached. These results are shown in Figure 5. We define the specifics of the experimental setup in its caption.

For $w = 1$, both variants of RTS do not show major differences. The major difference in the performance starts when $w \geq 2$ for all populations sizes. RTS without replacement is able to avoid the extinction of one subpopulation for smaller population sizes than its variant with replacement, i. e., while RTS without replacement is able to maintain individuals with fitness values up to 65 for small population sizes ($\mu = 2$) and small window sizes ($w \geq 2$), its variant with replacement requires greater window sizes ($w \geq 8$) for the same population size to achieve similar results. In general, both RTS variants have a better performance for large μ and for large w , which allows more individuals to participate in the tournaments. The main difference between both RTS variants is that RTS with replacement is not able to maintain populations on both branches of TWOMAX for a long time after initialisation for small μ and w . Takeover happens when the individuals have a fitness of around 50, which indicates that takeover happens close to initialisation. This makes sense since individuals on different branches may look similar to each other and with small w , unwanted replacements may happen.

RTS without replacement is more resistant to extinction after initialisation for small μ and w , but requires more fitness evaluations. This can be observed simply from the

E. Covantes Osuna and D. Sudholt

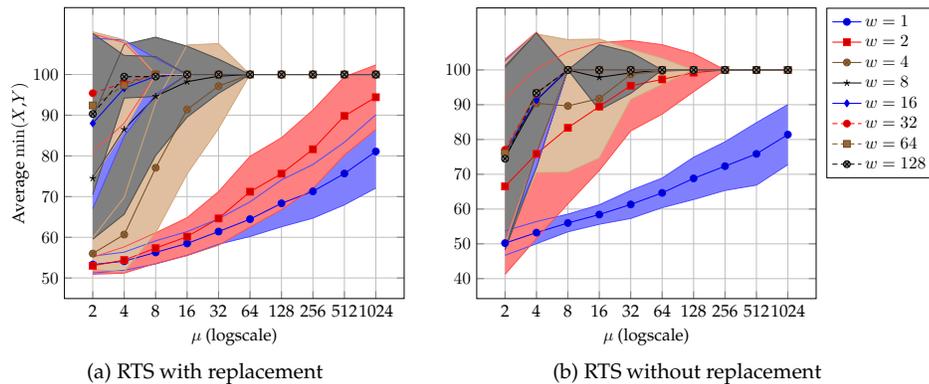


Figure 5: Average (pointed line) and Standard Deviation (shaded area) of minimum fitness reached between the best individual X found from the 0^n branch the best individual Y found from the 1^n branch ($\min(X, Y)$) measured among 100 runs at the time both optima were found on TWOMAX, the population has collapsed into one optimum on TWOMAX (the population consists of copies of just 0^n or 1^n) with no time budget for $n = 100$ with the $(\mu+1)$ EA with restricted tournament selection with and without replacement for $\mu \in \{2, 4, 8, \dots, 1024\}$, $w \in \{1, 2, 4, 8, \dots, 128\}$ and genotypic distance.

results on Figure 5 but specifically if we compare them with $w \geq 2$ and $\mu \geq 2$, RTS without replacement is able to reach fitness values above 65 compared to its variant with replacement.

6 Conclusion

We theoretically and empirically examined the behaviour of restricted tournament selection, embedded into a simple $(\mu+1)$ EA, on the bimodal function TWOMAX, where the goal is to find both optima. We rigorously proved that the performance of RTS varies a lot with the window size w . If w is large enough, $w \geq 2.5\mu \ln n$, then RTS behaves similarly to deterministic crowding. The probability of finding both optima in time $O(\mu n \log n)$ is close to $1 - 2^{-\mu+1}$, hence converging to 1 very quickly as μ grows. For small μ and w , if $\mu \leq n^{1/(w-1)}$ (and $w \geq 3$) then RTS likely fails to find both optima of TWOMAX. This even holds when the $(\mu+1)$ EA is allowed to start with the most promising population that does not yet contain both optima.

When selecting individuals for the tournament *without* replacement, the tournament becomes more diverse. While the positive result for the original RTS easily transfers with the more lax condition $w \geq \mu$, the population dynamics in the case of small w become more complicated. Experiments suggest that for small values of w typically one niche collapses to a single individual, and we proved that under certain conditions the algorithm is very slow at climbing up said branch. We conjecture a lower time bound of $\Omega((\mu^2/w) \cdot n \log n)$ for small values of w , where it is common for a single “lone” individual to evolve. This lower bound is by a factor of μ/w larger than the upper time bound for RTS with large w (Theorems 3 and 6). Experiments support this conjecture as for $n = 100$ the runtime increases drastically when w is very small. Table 2 showed that the average runtime is high when lone individuals emerged, and lone individuals typically emerged for very small w and small μ .

Our theoretical results cover small and large values for the window size w . It is still an open problem to theoretically analyse the population dynamics for both RTS variants for intermediate values for w . Our experiments indicate that RTS with and without replacement can optimise TWOMAX for smaller w than those required in Theorems 3 and 6. Showing refined upper bounds for smaller values of w that make this rigorous remains an open problem. Likewise, showing improved lower bounds for larger values of w than those given in Theorem 5, or proving the conjectured lower bound of $\Omega((\mu^2/w)n \log n)$ for RTS without replacement remain open problems. Note that proving lower bounds for population-based algorithms is a notoriously hard challenge even in the absence of diversity mechanisms, which may require new methods to be developed (see, e. g. Sutton and Witt (2019); Oliveto et al. (2020) for recent approaches in this direction).

Acknowledgments

The authors would like to thank the Consejo Nacional de Ciencia y Tecnología — CONACYT (the Mexican National Council for Science and Technology) for the financial support under the grant no. 409151 and registration no. 264342.

References

- Covantes Osuna, E. and Sudholt, D. (2018a). Empirical Analysis of Diversity-Preserving Mechanisms on Example Landscapes for Multimodal Optimisation. In *Proc. of PPSN XV*, pages 207–219. Springer.
- Covantes Osuna, E. and Sudholt, D. (2018b). Runtime Analysis of Probabilistic Crowding and Restricted Tournament Selection for Bimodal Optimisation. In *Proc. of GECCO '18*, pages 929–936. ACM.
- Covantes Osuna, E. and Sudholt, D. (2019). On the Runtime Analysis of the Clearing Diversity-Preserving Mechanism. *Evolutionary Computation Journal*, 27(3):403–433.
- Covantes Osuna, E. and Sudholt, D. (2020). Runtime Analysis of Crowding Mechanisms for Multimodal Optimization. *IEEE Transactions on Evolutionary Computation*, 24(3):581–592.
- Črepinšek, M., Liu, S.-H., and Mernik, M. (2013). Exploration and Exploitation in Evolutionary Algorithms: A Survey. *ACM Comput. Surv.*, 45(3):35:1–35:33.
- Doerr, B. (2020). Probabilistic Tools for the Analysis of Randomized Optimization Heuristics. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 1–87. Springer International Publishing.
- Doerr, B. and Goldberg, L. A. (2010). Drift Analysis with Tail Bounds. In *Proc. of PPSN XI*, pages 174–183. Springer.
- Doerr, B. and Goldberg, L. A. (2013). Adaptive drift analysis. *Algorithmica*, 65:224–250.
- Friedrich, T., Oliveto, P. S., Sudholt, D., and Witt, C. (2009). Analysis of Diversity-preserving Mechanisms for Global Exploration. *Evolutionary Computation*, 17(4):455–476.
- Galán, S. F. and Mengshoel, O. J. (2010). Generalized Crowding for Genetic Algorithms. In *Proc. of GECCO '10*, pages 775–782. ACM.

E. Covantes Osuna and D. Sudholt

- García-Hernández, L., Palomo-Romero, J. M., Salas-Morera, L., Arauzo-Azofra, A., and Pierreal, H. (2015). A novel hybrid evolutionary approach for capturing decision maker knowledge into the unequal area facility layout problem. *Expert Systems with Applications*, 42(10):4697–4708.
- García-Martínez, C., Lozano, M., and Rodríguez-Díaz, F. (2012). A simulated annealing method based on a specialised evolutionary algorithm. *Applied Soft Computing*, 12(2):573–588.
- Glibovets, N. N. and Gulayeva, N. M. (2013). A Review of Niching Genetic Algorithms for Multimodal Function Optimization. *Cybernetics and Systems Analysis*, 49(6):815–820.
- Harik, G. R. (1995). Finding Multimodal Solutions Using Restricted Tournament Selection. In *Proc. of the 6th ICGA*, pages 24–31. Morgan Kaufmann Publishers Inc.
- Levin, D., Peres, Y., and Wilmer, E. (2009). *Markov Chains and Mixing Times*. AMS.
- Mengsheol, O. and Goldberg, D. (1999). Probabilistic Crowding: Deterministic Crowding with Probabilistic Replacement. In *Proc. of GECCO '99*, pages 409–416.
- Oliveto, P. S., Sudholt, D., and Witt, C. (2020). A Tight Lower Bound on the Expected Runtime of Standard Steady State Genetic Algorithms. In *Proc. GECCO '20*, pages 1323–1331. ACM.
- Oliveto, P. S., Sudholt, D., and Zarges, C. (2019). On the benefits and risks of using fitness sharing for multimodal optimisation. *Theoretical Computer Science*, 773:53–70.
- Qu, B. Y. and Suganthan, P. N. (2010). Novel multimodal problems and differential evolution with ensemble of restricted tournament selection. In *Proc. of IEEE CEC*, pages 1–7.
- Sareni, B. and Krahenbuhl, L. (1998). Fitness sharing and niching methods revisited. *IEEE Transactions on Evolutionary Computation*, 2(3):97–106.
- Shir, O. M. (2012). Niching in Evolutionary Algorithms. In *Handbook of Natural Computing*, pages 1035–1069. Springer.
- Singh, G. and Deb, K. (2006). Comparison of Multi-modal Optimization Algorithms Based on Evolutionary Algorithms. In *Proc. of GECCO '06*, pages 1305–1312. ACM.
- Squillero, G. and Tonda, A. (2016). Divergence of character and premature convergence: A survey of methodologies for promoting diversity in evolutionary optimization. *Information Sciences*, 329:782–799.
- Sudholt, D. (2020). The Benefits of Population Diversity in Evolutionary Algorithms: A Survey of Rigorous Runtime Analyses. In *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, pages 359–404. Springer International Publishing.
- Sutton, A. M. and Witt, C. (2019). Lower Bounds on the Runtime of Crossover-Based Algorithms via Decoupling and Family Graphs. In *Proc. GECCO '19*, pages 1515–1522. ACM.
- Tsafarakis, S. (2016). Redesigning product lines in a period of economic crisis: a hybrid simulated annealing algorithm with crossover. *Annals of Operations Research*, 247(2):617–633.