

On the Impact of the Mutation-Selection Balance on the Runtime of Evolutionary Algorithms

Per Kristian Lehre

The Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom
p.k.lehre@cs.bham.ac.uk

Xin Yao

The Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom
x.yao@cs.bham.ac.uk

ABSTRACT

The interplay between the mutation operator and the selection mechanism plays a fundamental role in the behaviour of evolutionary algorithms. However, this interplay is still not completely understood. This paper presents a rigorous runtime analysis of a non-elitistic population based evolutionary algorithm that uses the linear ranking selection mechanism. The analysis focuses on how the balance between parameter η controlling the selection pressure in linear ranking selection, and parameter χ controlling the bit-wise mutation rate impacts the expected runtime.

The results point out situations where a correct balance between selection pressure and mutation rate is essential for finding the optimal solution in polynomial time. In particular, it is shown that there exist fitness functions which under a certain assumption can be solved in polynomial time if the ratio between parameters η and χ is appropriately tuned to the problem instance class, but where a small change in this ratio can increase the runtime exponentially. Furthermore, it is shown that the appropriate parameter choice depends on the characteristics of the fitness function. Hence there does in general not exist a problem-independent optimal balance between mutation rate and selection pressure.

The results are obtained using new techniques based on branching processes.

Categories and Subject Descriptors

F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems—*Computations on discrete structures*; G.3 [Probability and Statistics]: Probabilistic Algorithms

General Terms

Algorithms, Theory

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

FOGA '09, January 9–11, 2009, Orlando, Florida, USA.
Copyright 2009 ACM 978-1-60558-414-0/09/01 ...\$5.00.

Keywords

Evolutionary algorithms, runtime analysis, branching processes

1. INTRODUCTION

Evolutionary algorithms (EAs) have been applied successfully to many optimisation problems [14]. However, despite several decades of research, many fundamental questions about their behaviour remain open. One of the central questions regarding EAs is to understand the interplay between the selection mechanism and the genetic operators. Several authors have suggested that EAs must find a balance between maintaining a sufficiently diverse population to explore new parts of the search space, and at the same time exploit the currently best found solutions by steering the search in this direction [6, 16, 7].

Much research has therefore focused on finding measures to quantify the selection pressure in selection mechanisms, and subsequently on investigating how EA parameters influence these measures [7, 1, 2, 15, 2, 3]. One such measure, called the *take-over time*, considers the behaviour of an evolutionary process consisting only of the selection step, and no crossover or mutation operators [7, 1]. Subsequent populations are produced by selecting individuals from the previous generation, keeping at least one copy of the fittest individual. Hence, the population will after a certain number of generations only contain those individuals that were fittest in the initial population, and this time is called the take-over time. A short take-over time corresponds to a high selection pressure. Other measures of selection pressure consider properties of the distribution of fitness values in a population that is obtained by a single application of the selection mechanism to a population with normally distributed fitness values [2]. One of these properties is the *selection intensity*, which is the difference between the average population fitness before and after selection [15]. Other properties include *loss of diversity* [2, 12] and higher order cumulants of the fitness distribution [3].

To completely understand the role of selection mechanisms, it is necessary to also take into account their interplay with the genetic operators. There exists few rigorous studies of selection mechanisms when used in combination with genetic operators. Happ et al. analysed variants of the RLS and (1+1) EA that use fitness proportionate selection, showing that both these algorithms have exponential

runtime on the class of linear functions [9]. However, the algorithms considered only use a single individual, so it is difficult to draw any conclusion regarding population-based algorithms. Witt analysed a population-based algorithm with fitness proportionate selection, however with the objective to study the role of populations [18]. Recently, Chen et al. have analysed the runtime of the $(N+N)$ Evolutionary Algorithm using either truncation selection, linear ranking selection or binary tournament selection on the LEADINGONES and ONEMAX fitness functions [4]. They show that the expected runtime on these fitness functions is the same for all three selection mechanisms. These results do not show how the balance between selection pressure and mutation rate impacts the runtime. It is worth noting that the considered algorithm in their work is elitistic, i.e. the best individual in every generation is always copied to the next generation.

This paper analyses rigorously a non-elitistic, population based EA that uses linear ranking selection and bitwise mutation. The main contributions are an analysis of situations where the mutation-selection balance has an exponentially large impact on the expected runtime, and new techniques based on branching processes for analysing non-elitistic population based EAs.

1.1 Notation and Preliminaries

The following notation will be used in the rest of this paper. The length of a bitstring x is denoted $\ell(x)$. The i th bit, $1 \leq i \leq \ell(x)$, of a bitstring x is denoted x_i . The concatenation of two bitstrings x and y is denoted by $x \cdot y$, and sometimes xy . Given a bitstring x , the notation $x[i, j]$, where $1 \leq i < j \leq \ell(x)$, denotes the substring $x_i x_{i+1} \cdots x_j$. For any bitstring x , define $\|x\| := \sum_{i=1}^{\ell(x)} x_i / \ell(x)$, i.e. the fraction of 1-bits in the bitstring.

In contrast to classical algorithms, the runtime of evolutionary algorithms is usually measured in terms of the number of evaluations of the fitness function, and not the number of basic operations.

Definition 1 (RUNTIME [5, 10]). *Given a class \mathcal{F} of fitness functions $f_i : S_i \rightarrow \mathbb{R}$, the runtime $T_{A, \mathcal{F}}(n)$ of a search algorithm A is defined as*

$$T_{A, \mathcal{F}}(n) := \max\{T_{A, f} \mid f \in \mathcal{F}_n\},$$

where \mathcal{F}_n is the subset of functions in \mathcal{F} with instance size n , and $T_{A, f}$ is the number of times algorithm A evaluates the cost function f until the optimal value of f is evaluated for the first time.

The variable name τ will be used to denote the runtime in terms of number of generations of the EA. Given a population size λ , this variable is related to the runtime T by $\lambda(\tau - 1) \leq T \leq \lambda\tau$.

2. DEFINITIONS

2.1 Linear Ranking Selection

In ranking selection, individuals are selected according to their fitness rank in the population. A ranking selection mechanism is uniquely defined by the probabilities p_i of selecting an individual ranked i , for all ranks i [2]. For mathematical convenience, an alternative definition due to Goldberg and Deb [7] is adopted, in which a function $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ is considered a ranking function if it satisfies the following three properties

1. $\alpha(x) \in \mathbb{R}$ for $x \in [0, 1]$,
2. $\alpha(x) \geq 0$, and
3. $\int_0^1 \alpha(y) dy = 1$.

Individuals are ranked from 0 to 1, with the best individual ranked 0, and the worst individual ranked 1. For a given ranking function α , the integral $\beta(x, y) := \int_x^y \alpha(z) dz$ gives the probability of selecting an individual with rank between x and y . By defining the linearly decreasing ranking function $\alpha(x) := \eta - c \cdot x$, where η and c are parameters, one obtains *linear ranking selection*. The ranking function properties imply that $\eta \geq c \geq 0$, and $c = 2 \cdot (\eta - 1)$. Hence, for linear ranking selection, we have

$$\alpha(x) := \eta \cdot (1 - 2x) + 2x, \text{ and} \quad (1)$$

$$\beta(x) := \beta(0, x) = x \cdot (\eta \cdot (1 - x) + x). \quad (2)$$

Given a fixed population size λ , the selection pressure, measured in terms of take-over time, is uniquely given by and monotonically decreasing in parameter η [7]. The weakest selection pressure is obtained for $\eta = 1$, where selection is uniform over the population, and the highest selection pressure is obtained for $\eta = 2$.

2.2 Evolutionary Algorithm

1 Linear Ranking EA

- 1: $t \leftarrow 0$.
 - 2: **for** $i = 1$ to λ **do**
 - 3: Sample x uniformly at random from $\{0, 1\}^n$.
 - 4: $P_0(i) \leftarrow x$.
 - 5: **end for**
 - 6: **repeat**
 - 7: Sort P_t according to fitness f , such that
 $f(P_t(1)) \geq f(P_t(2)) \geq \cdots \geq f(P_t(\lambda))$.
 - 8: **for** $i = 1$ to λ **do**
 - 9: Sample r in $\{1, \dots, \lambda\}$ with $\Pr[r \leq \gamma\lambda] = \beta(\gamma)$.
 - 10: $P_{t+1}(i) \leftarrow P_t(r)$.
 - 11: Flip each bit position in $P_{t+1}(i)$ with prob. χ/n .
 - 12: **end for**
 - 13: $t \leftarrow t + 1$.
 - 14: **until** termination condition met.
-

We consider a population-based non-elitistic evolutionary algorithm which uses linear ranking as selection mechanism. The crossover operator will not be considered in this paper. The pseudo-code of the algorithm is given above. After sampling the initial population P_0 at random in lines 1 to 5, the algorithm enters its main loop where the current population P_t in generation t is sorted according to fitness, then the next population P_{t+1} is generated by independently selecting (line 9) and mutating (line 10) individuals from the previous population P_t . The analysis of the algorithm is based on the assumption that parameter χ is a constant with respect to n .

Linear ranking selection is indicated in line 9, where for a given selection pressure η , the cumulative probability of sampling individuals with rank less than $\gamma \cdot \lambda$ is $\beta(\gamma)$. It can be seen from the definition of the functions α and β , that the upper bound $\beta(\gamma, \gamma + \delta) \leq \delta \cdot \alpha(\gamma)$, holds for any $\gamma, \delta > 0$ where $\gamma + \delta \leq 1$. Hence, the expected number of times a uniformly chosen individual ranked between $\gamma\lambda$ and

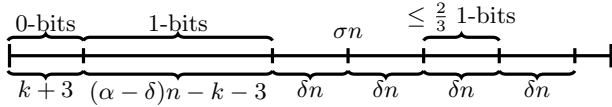


Figure 1: Illustration of optimal search points.

$(\gamma + \delta)\lambda$ is selected during one generation is upper bounded by $(\lambda/\delta\lambda) \cdot \beta(\gamma, \gamma + \delta) \leq \alpha(\gamma)$. We leave the implementation details of the sampling strategy unspecified, and assume that the EA has access to some sampling mechanism which draws samples perfectly according to β .

2.3 Fitness Function

Definition 2. For any constants σ , $0 < \sigma < 1$, and integer $k \geq 1$, define the fitness function parameterised by σ and k as

$$\text{SELPRES}_{\sigma,k}(x) := \begin{cases} 2n & \text{if } x \in X_{\sigma}^*, \text{ and} \\ \sum_{i=1}^n \prod_{j=1}^i x_j & \text{otherwise,} \end{cases}$$

where the set of optimal solutions X_{σ}^* is defined to contain all bitstrings $x \in \{0, 1\}^n$ satisfying

$$\begin{aligned} \|x[1, k+3]\| &= 0, \\ \|x[k+4, (\sigma - \delta)n - 1]\| &= 1, \text{ and} \\ \|x[(\sigma + \delta)n, (\sigma + 2\delta)n - 1]\| &\leq 2/3, \end{aligned}$$

where $\delta > 0$ is any arbitrarily small constant.

Except for the set of globally optimal solutions X_{σ}^* , the fitness function takes the same values as the well known LEADINGONES fitness function, i.e. the number of leading 1-bits in the bitstring. The form of the optimal search points, which is illustrated in Figure 1, depends on the three problem parameters σ, k and δ . The δ -parameter is needed for technical reasons and can be set arbitrarily close to 0. Hence, the globally optimal solutions have approximately σn leading 1-bits, except for $k+3$ leading 0-bits. In addition, globally optimal search points must have a short interval after the first σn bits which does not contain too many 1-bits.

3. MAIN RESULT

Theorem 1. Let T be the runtime of the Linear Ranking EA with population size $n \leq \lambda \leq n^k$ using linear ranking selection with a constant selection pressure of $\eta, 1 < \eta \leq 2$, and bit-wise mutation rate χ/n , with constant $\chi > 0$, on fitness function $\text{SELPRES}_{\sigma,k}$ with parameter $\sigma, 0 < \sigma < 1$. Then for any constant $\epsilon > 0$,

- If $\eta > 2 \exp(\chi(\sigma + 3\delta + \epsilon)) - 1 + \delta$, then

$$\mathbf{E}[T] = e^{\Omega(n)}.$$

- If $\eta = \exp(\chi\sigma)$ and no individual reaches more than $(\sigma + \delta)n$ leading 1-bits, then

$$\mathbf{E}[T] = O(n^{k+3}).$$

PROOF. The theorem follows from Theorems 5 and 6. \square

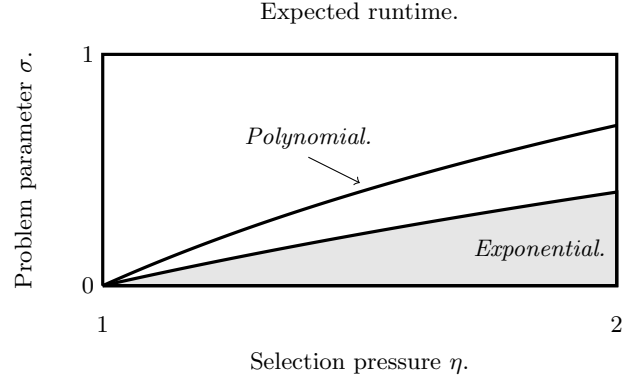


Figure 2: Illustration of Theorem 1 with fixed mutation rate $\chi = 1$, indicating expected runtime of the EA on $\text{SELPRES}_{\sigma,k}$ as a function of selection pressure η (horizontal axis) and problem parameter σ (vertical axis).

Theorem 1 is illustrated in Figure 2 for mutation rate $\chi = 1$. Each point in the gray area indicates that for the corresponding values of selection pressure η and problem parameter σ , the EA has expected exponential runtime (i.e. is highly inefficient). The thick line indicates values of σ and η where the expected runtime of the EA is a small polynomial (i.e. is highly efficient), under the assumption that no individual reaches more than $(\sigma + \delta)n$ leading 1-bits. The expected runtime in the white regions is not analysed.

4. RUNTIME ANALYSIS

This section gives the proofs of Theorem 1. The analysis is conceptually divided into two parts. In Sections 4.1 and 4.2, the behaviour of the main “core” of the population is analysed, showing that the population enters an equilibrium state. This analysis is sufficient to prove the polynomial upper bound in Theorem 1. Sections 4.3 and 4.4 analyse the behaviour of the “stray” individuals that sometimes move away from the core of the population. This analysis is necessary to prove the exponential lower bound in Theorem 1.

4.1 Population Equilibrium

As long as the global optimum has not been found, the population is evolving with respect to the number of leading 1-bits. In the following, we will prove that the population eventually reaches an equilibrium state in which the population makes no progress with respect to the number of leading 1-bits.

The population equilibrium can be explained informally as follows. On one hand, the selection mechanism increases the number of individuals in the population that have a relatively high number of leading 1-bits. On the other hand, the mutation operator may flip one of the leading 1-bits, and the probability of doing so clearly increases by the number of leading 1-bits in the individual. Hence, the selection mechanism causes an influx of individuals with a high number of leading 1-bits, and the mutation causes an efflux of individuals with a high number of leading 1-bits. At a certain point, the influx and efflux reach a balance which is described in the field of population genetics as mutation selection balance.

Our first goal will be to describe the population when it

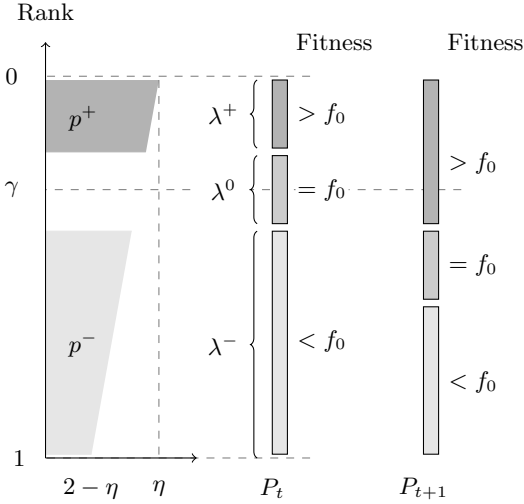


Figure 3: Impact of one generation of selection and mutation from the point of view of the γ -ranked individual in population P_t .

is in the equilibrium state. This is done rigorously by considering each generation as a sequence of λ Bernoulli trials, where each trial consists of selecting an individual from the population and then mutating that individual. Each trial has a certain probability of being successful in a sense that will be described later, and the progress of the population depends on the sum of successful trials, i.e. the population progress is a function of a certain Bernoulli process.

4.1.1 Ranking Selection as a Bernoulli Process

We will associate a Bernoulli process with the selection step in any given generation of the non-elitistic EA, similar to Chen et al [4]. For notational convenience, the individual that has rank $\gamma \cdot \lambda$ in a given population, will be called the γ -ranked individual of that population. For any constant $\gamma, 0 < \gamma < 1$, assume that the γ -ranked individual has $f_0 := \xi n$ leading 1-bits for some constant ξ . As illustrated in Figure 3, the population can be partitioned into three groups of individuals: λ^+ individuals with fitness higher than f_0 , λ^0 individuals with fitness equal to f_0 , and λ^- individuals with fitness less than f_0 . Clearly, $\lambda^+ + \lambda^0 + \lambda^- = \lambda$, and $0 \leq \lambda^+ < \gamma \cdot \lambda$.

Theorem 2. For any constant $\gamma, 0 < \gamma < 1$, let ξn be the number of leading 1-bits in the γ -ranked individual of a population which does not contain an optimal solution. Then for any constant $\delta > 0$,

1. if $\xi < \ln(\beta(\gamma)/\gamma)/\chi - \delta$, then the probability that the γ -ranked individual in the next generation has at least ξn leading 1-bits is $1 - e^{-\Omega(\lambda)}$, and
2. if $\xi > \ln(\beta(\gamma)/\gamma)/\chi + \delta$ then the probability that the γ -ranked individual in the next generation has at most ξn leading 1-bits is $1 - e^{-\Omega(\lambda)}$,

where $\beta(\gamma)$ is as given in Eq. (2).

PROOF. For the first part of the theorem, we consider each iteration of the selection mechanism a Bernoulli trial where a trial is successful if the following event occurs:

\mathcal{E}_1^+ : An individual with at least ξn leading 1-bits is selected, and none of the initial ξn bits are flipped.

Let random variable X denote the number of successful trials. Notice that the event $X \geq \gamma \cdot \lambda$ implies that the γ -ranked individual in the next generation has at least ξn leading 1-bits. The assumption $\xi < \ln(\beta(\gamma)/\gamma)/\chi - \delta$ implies that

$$\begin{aligned} \mathbf{E}[X] &= \lambda \cdot \Pr[\mathcal{E}_1^+] \\ &\geq \lambda \cdot \beta(\gamma) \cdot (1 - \chi/n) \cdot (1 - \chi/n)^{\xi n - 1} \\ &\geq \lambda \cdot \beta(\gamma) \cdot (1 - \chi/n) \cdot e^{-\xi \chi} \\ &\geq \gamma \cdot \lambda \cdot (1 - \chi/n) \cdot e^{x\delta} \\ &\geq (1 + \chi\delta) \cdot \gamma \cdot \lambda \cdot (1 - \chi/n). \end{aligned}$$

For sufficiently large n , a Chernoff bound [13] therefore implies that the probability that the number of successes is less than $\gamma \cdot \lambda$ is $e^{-\Omega(\lambda)}$.

For the second part of the theorem, we define a trial successful if one of the following two events occurs:

\mathcal{E}_2^+ : An individual with at least $\xi n + 1$ leading 1-bits is selected, and none of the initial $\xi n + 1$ bits are flipped.

\mathcal{E}_2^- : An individual with less than $\xi n + 1$ leading 1-bits is selected, and the mutation of this individual creates an individual with at least $\xi n + 1$ leading 1-bits.

Let random variable Y denote the number of successful trials. Notice that the event $Y < \gamma \cdot \lambda$ implies that the γ -ranked individual in the next generation has no more than ξn leading 1-bits. Furthermore, since the γ -ranked individual in the current generation has exactly ξn leading 1-bits, less than $\gamma \cdot \lambda$ individuals have more than ξn leading 1-bits. The probability of the event \mathcal{E}_2^+ is therefore bounded by

$$\Pr[\mathcal{E}_2^+] \leq \beta(\gamma) \cdot (1 - \chi/n)^{\xi n + 1} \leq \beta(\gamma)/e^{\xi \chi}.$$

If the selected individual has $k \geq 1$ 0-bits within the first $\xi n + 1$ bit positions, then the probability of mutating this individual into an individual with at least $\xi n + 1$ leading 1-bits, and hence also the probability of event \mathcal{E}_2^- , is bounded from above by

$$\Pr[\mathcal{E}_2^-] \leq (1 - \chi/n)^{\xi n + 1 - k} \cdot (\chi/n)^k \leq \chi/n e^{\xi \chi}.$$

The assumption $\xi \geq \ln(\beta(\gamma)/\gamma)/\chi + \delta$ then implies that for any constant $\delta', 0 < \delta' < 1 - e^{-\delta \xi} < 1$,

$$\begin{aligned} \mathbf{E}[Y] &= \lambda \cdot (\Pr[\mathcal{E}_2^+] + \Pr[\mathcal{E}_2^-]) \\ &\leq \lambda \cdot (\beta(\gamma) + \chi/n) \cdot e^{-\xi \chi} \\ &\leq \gamma \cdot \lambda \cdot (1 + \chi/n\beta(\gamma)) \cdot e^{-\chi\delta} \\ &\leq (1 - \delta') \cdot \gamma \cdot \lambda \cdot (1 + \chi/n\beta(\gamma)). \end{aligned}$$

For sufficiently large n , a Chernoff bound therefore implies that the probability that the number of successes is at least $\gamma \cdot \lambda$ is $e^{-\Omega(\lambda)}$. \square

In the following, we will say that the γ -ranked individual x is in the equilibrium position if the number of leading 1-bits in x is higher than $(\xi - \delta)n$ and smaller than $(\xi + \delta)n$, where $\xi = \ln(\beta(\gamma)/\gamma)/\chi$.

4.1.2 Drift Analysis in two Dimensions

Theorem 2 states that when the population reaches a certain area of the search space, the progress of the population will halt and the EA enters an equilibrium state. Our

next goal is to calculate the expected time until the EA enters the equilibrium state. More precisely, for any constants $\gamma, 0 < \gamma < 1$ and $\delta > 0$, we would like to bound the expected number of generations until the fitness f_0 of the γ -ranked individual becomes at least $n \cdot (\ln(\beta(\gamma)/\gamma)/\chi - \delta)$. Although the fitness f_0 will have a tendency to drift towards higher values, it is necessary to take into account that the fitness can in general both decrease and increase according to stochastic fluctuations.

Drift analysis has proven to be a powerful mathematical technique to analyse such stochastically fluctuating processes [10]. Given a distance measure (sometimes called potential function) from any search point to the optimum, one estimates the expected drift Δ towards the optimum in one generation, and bounds the expected time to overcome a distance of $b(n)$ by $b(n)/\Delta$.

However, in our case, a direct application of drift analysis with respect to f_0 will give poor bounds, because the expected drift of f_0 depends on the value of a second variable λ^+ . The probability of increasing the fitness of the γ -ranked individual is low when the number of individuals in the population with higher fitness, i.e. λ^+ , is low. However, it is still likely that the sum $\lambda^0 + \lambda^+$ will increase, thus increasing the number of good individuals in the population.

Several researchers have discussed this form of alternating behaviour in population-based EAs [17, 4]. Witt shows that by taking into account replication of good individuals, one can improve on trivial upper runtime bounds for the $(\mu+1)$ EA, e.g. from $O(\mu n^2)$ on the LEADINGONES problem into $O(\mu n \log n + n^2)$ [17]. Chen et al. describes a similar situation in the case of an elitistic EA, which goes through a sequence of two-stage phases, where the first stage is characterised by accumulation of leading individuals, and the second stage is characterised by acquiring better individuals [4].

Generalised to the non-elitistic EA described here, this corresponds to first accumulation of λ^+ -individuals, until one eventually gains more than $\gamma\lambda$ individuals with fitness higher than f_0 . In the worst case, when $\lambda^+ = 0$, one expects that f_0 has a small positive drift. However, when λ^+ is high, there is a high drift. When the fitness is increased, the value of λ^+ is likely to decrease. To take into account this mutual dependency between λ^+ and f_0 , we apply drift analysis in conceptually two dimensions, finding the expected drift of both f_0 and λ^+ . The drift analysis applies the following simple property of function β which follows directly from the definition in Eq. (2).

Lemma 1. *The function β defined in Eq. (2) satisfies $\beta(\gamma/l)/\beta(\gamma) \geq 1/l$, for all $\gamma, 0 < \gamma < 1$, and $l \geq 1$.*

The following theorem is analogous to Theorem 3, and shows that if the γ -ranked individual in a given population is below the equilibrium position, then the equilibrium position will be reached within expected $O(n^2)$ generations.

Theorem 3. *Let γ and δ be any constants with $0 < \gamma < 1$ and $\delta > 0$. The expected number of function evaluations until the γ -ranked individual of the Linear Ranking EA with population size $\lambda \geq c \ln n$, for some constant $c > 0$, attains at least $n(\ln(\beta(\gamma)/\gamma)/\chi - \delta)$ leading 1-bits or the optimum is reached, is $O(\lambda n^2)$.*

PROOF. We consider the drift according to the potential function $p(X_t) := h(X_t) + \lambda \cdot g(X_t)$, which is composed of a

horizontal component g , and a vertical component h , defined as

$$\begin{aligned} g(X_t) &:= n - \text{LEADINGONES}(x_{(\gamma)}), \\ h(X_t) &:= \gamma \cdot \lambda - |\{y \in P_t \mid f(y) > f(x_{(\gamma)})\}| \end{aligned}$$

where $x_{(\gamma)}$ is the γ -ranked individual in population X_t . The horizontal $\Delta_{x,t}$ and vertical $\Delta_{y,t}$ drift in generation t is $\Delta_{x,t} := g(X_t) - g(X_{t+1})$, and $\Delta_{y,t} := h(X_t) - h(X_{t+1})$. The horizontal and vertical drift will be bounded independently in the following two cases,

- 1) $0 \leq \lambda_t^+ \leq \gamma\lambda/l$, and
- 2) $\gamma\lambda/l < \lambda_t^+$,

where l is a constant that will be specified later,

Assume that the γ -ranked individual has ξn leading 1-bits, where $\xi < \ln(\beta(\gamma)/\gamma)/\chi - \delta$. The horizontal distance cannot increase by more than n , so by Theorem 2, the expected horizontal drift in both cases is at least $\Delta_{x,t} \geq -ne^{-\Omega(\lambda)}$.

We now bound the horizontal drift Δ_x for Case 2. Let the random variable S_t denote the number of selection steps in which an individual with fitness strictly higher than $f_0 = f(x_{(\gamma)})$ is selected, and none of the leading ξn bits are flipped. The expectation of S_t is bounded by

$$\begin{aligned} \mathbf{E}[S_t] &\geq \lambda \cdot \beta(\gamma/l) \cdot e^{-\xi x} \cdot (1 - \chi/n) \\ &\geq \gamma\lambda \cdot (1 + \chi\delta) \cdot \frac{\beta(\gamma/l)}{\beta(\gamma)} \cdot (1 - \chi/n) \\ &\geq \gamma\lambda \cdot \frac{(1 + \chi\delta)}{l} \cdot (1 - \chi/n). \end{aligned}$$

By defining $l := (1 + \chi\delta/2)$, there exists a constant $\delta' > 0$ such that for sufficiently large n , we have $\mathbf{E}[S_t] \geq (1 + \delta') \cdot \gamma\lambda$. Hence, by a Chernoff bound, with probability $1 - e^{-\Omega(\lambda)}$, the number S_t of such selection steps is at least $\gamma\lambda$, and hence $\Delta_{t,x} \geq 1$. The horizontal drift in Case 2 is therefore $\Delta_x \geq 1 \cdot (1 - e^{-\Omega(\lambda)}) - n \cdot e^{-\Omega(\lambda)}$.

We now bound the vertical drift Δ_y for Case 1. In order to generate a λ^+ -individual in a selection step, it is sufficient that a λ^+ -individual is selected and none of the leading $\xi n + 1$ 1-bits is flipped. Assuming that $\lambda_t^+ = \gamma\lambda/m$ for some constant $m > 1$, the expected number of such events is at least

$$\begin{aligned} &\lambda \cdot \beta(\gamma/m) \cdot e^{-\xi x} \cdot (1 - \chi/n) \\ &\geq \gamma\lambda \cdot \frac{\beta(\gamma/m)}{\beta(\gamma)} \cdot (1 + \chi\delta) \cdot (1 - \chi/n) \\ &\geq (\lambda\gamma/m) \cdot (1 + \chi\delta) \cdot (1 - \chi/n). \end{aligned}$$

Hence, for sufficiently large n , this is at least λ_t^+ , and the expected vertical drift is at least positive. In addition, a λ^+ -individual can be created by selecting a λ^0 -individual, and flipping the first 0-bit and no other bits. The expected number of such events is at least $\lambda \cdot \beta(\gamma/l, \gamma) \cdot e^{-\xi x} \cdot \chi/n = \Omega(\lambda/n)$. Hence, the expected vertical drift in Case 1 is $\Omega(\lambda/n)$. Finally, for Case 2, we use the trivial lower bound $\Delta_y \geq -\gamma\lambda$.

The horizontal and vertical drift is now added into a combined drift $\Delta := \Delta_y + \lambda \cdot \Delta_x$, which in the two cases is bounded by

- 1) $\Delta = \Omega(\lambda/n) - \lambda \cdot n \cdot e^{-\Omega(\lambda)}$, and
- 2) $\Delta = -\gamma \cdot \lambda + \lambda \cdot (1 - e^{-\Omega(\lambda)}) - \lambda \cdot n \cdot e^{-\Omega(\lambda)}$.

Given a population size $\lambda \geq c \ln n$, for a sufficiently large constant c , the combined drift Δ is therefore in both cases bounded from below by $\Omega(\lambda/n)$. The maximal distance is $b(n) \leq (n + \gamma) \cdot \lambda$, hence, the expected number of function evaluations T until the γ -ranked individual attains at least $n(\ln(\beta(\gamma)/\gamma)/\chi - \delta)$ leading 1-bits is no more than $\mathbf{E}[T] \leq \lambda \cdot b(n)/\Delta = O(\lambda n^2)$. \square

The following theorem shows that if the γ -ranked individual in a given population is above the equilibrium position, then the equilibrium position will be reached within expected $O(n)$ generations.

Theorem 4. *Let γ be any constant $0 < \gamma < 1$. If the γ -ranked individual has more than $(\xi + \delta + \epsilon) \cdot n$ leading 1-bits, with $\xi := \ln(\beta(\gamma)/\gamma)/\chi$ for any constants $\delta, \epsilon > 0$, then the expected number of generations until the γ -ranked individual has no more than $(\xi + \delta) \cdot n$ leading 1-bits or the optimum is reached, is $O(n)$.*

PROOF. We consider the drift according to a potential function $p(P_t) := h(P_t) + (\lambda + 1) \cdot g(P_t)$ that has a horizontal component g , and a vertical component h defined as

$$g(P_t) := \text{LEADINGONES}(x_{(\gamma)}) - (\xi + \delta) \cdot n,$$

$$h(P_t) := |\{y \in P_t \mid f(y) \geq f(x_{(\gamma)})\}|.$$

The vertical distance relates to the number of λ^- -individuals by $\lambda^- = \lambda - h(P_t)$, implying that if the number of λ^- -individuals increases, then the vertical distance decreases. Define $\gamma^0 := h(X_t)/\lambda$. A λ^- -individual is produced in one selection step if one of the two events \mathcal{E}^- and \mathcal{E}^+ occurs:

Event \mathcal{E}^- occurs when a λ^- -individual is selected, and one of the 0-bits in the interval from 1 to $\text{LEADINGONES}(x_{(\gamma)})$ is not flipped. At least one such 0-bit must exist in any λ^0 -individual. This event happens with probability at least $\Pr[\mathcal{E}^-] = (1 - \beta(\gamma^0)) \cdot (1 - \chi/n) \geq 1 - \chi/n - \beta(\gamma^0)$.

Event \mathcal{E}^+ occurs when a λ^+ - or λ^0 -individual is selected, and at least one of the leading $(\xi + \delta) \cdot n$ leading 1-bits is flipped. Noting that $\gamma \leq \gamma^0$ implies $\gamma \cdot \beta(\gamma^0)/\beta(\gamma) \leq \gamma^0$, the probability of this event can be bounded by

$$\begin{aligned} \Pr[\mathcal{E}^+] &\geq \beta(\gamma^0) \cdot \chi \cdot (\xi + \delta) \\ &= \beta(\gamma^0) \cdot (\ln(\beta(\gamma)/\gamma) + \chi\delta) \\ &\geq \beta(\gamma^0) \cdot (1 - \gamma/\beta(\gamma) + \chi\delta) \\ &\geq \beta(\gamma^0) - \gamma^0 + \beta(\gamma^0) \cdot \chi\delta. \end{aligned}$$

We now distinguish between two cases. In the first case, the number of λ^0 -individuals created during one generation is less than $(1 - \gamma) \cdot \lambda$. In this case, we bound the horizontal drift to $\Delta_x \geq -n \cdot e^{-\lambda}$ using Theorem 2. For sufficiently large n , the expected vertical drift is in this case bounded by

$$\begin{aligned} \Delta_y &\geq g(P_t) - \lambda \cdot (1 - \Pr[\mathcal{E}^-] - \Pr[\mathcal{E}^+]) \\ &\geq \lambda \cdot \gamma^0 - \lambda \cdot (\gamma^0 - \beta(\gamma^0) \cdot \chi\delta + \chi/n) = \Omega(\lambda). \end{aligned}$$

In the second case, the number of λ^0 -individuals produced during one generation is $(1 - \gamma) \cdot \lambda$ or larger, and the number of leading 1-bits in the γ -ranked individual must therefore have decreased. The vertical and horizontal drift can in this case be bounded by $\Delta_x \geq 1$ and $\Delta_y \geq -\lambda$.

Combining the horizontal and vertical drift $\Delta := \Delta_y + (\lambda + 1) \cdot \Delta_x$ now gives that the drift is bounded by $\Delta = \Omega(\lambda)$

in both cases. The maximal distance is $b(n) \leq (\lambda + 1) \cdot n + \lambda$, hence the expected number of function evaluations T until the γ -ranked individual has no more than $(\xi + \delta) \cdot n$ leading 1-bits is no more than $\mathbf{E}[T] \leq \lambda \cdot b(n)/\Delta = O(\lambda \cdot n)$. \square

4.2 Mutation-Selection Balance

In the previous section, it was shown that the population reaches an equilibrium state in $O(\lambda n^2)$ function evaluations in expectation. Furthermore, the position of the equilibrium state is given by the selection pressure η and the mutation rate χ . By choosing appropriate values for the parameters η and χ , one can ensure that the equilibrium position occurs close to the global optimum that is given by the problem parameter σ . Assuming that the best individual will not reach too many leading 1-bits, it is straightforward to prove that an optimal solution will be found in expected polynomial time, implying a polynomial upper bound on the expected runtime of the Linear Ranking EA on $\text{SELPRES}_{\sigma,k}$.

Theorem 5. *If no individual reaches more than $(\sigma + \delta)n$ leading 1-bits, then the expected runtime of the Linear Ranking EA on fitness function $\text{SELPRES}_{\sigma,k}$ when using population size $c \ln n < \lambda \leq n^k$, for some constant $c > 2$, and selection pressure η and bit-wise mutation rate χ/n satisfying $\eta = \exp(\sigma\chi)$ is $O(n^{k+3})$.*

PROOF. Let $\gamma > 0$ be any constant where $\ln(\beta(\gamma)/\gamma)/\chi > \sigma - \delta$. Let \mathcal{E} be the event that all individuals ranked between 0 and γ have at least $(\sigma - \delta)n$ leading 1-bits and at most $(\sigma + \delta)n$ leading 1-bits, and at most $2n\delta/3$ 1-bits in the interval from $n(\sigma + \delta)$ to $n(\sigma + 2\delta)$. Let random variable τ_c be the number of generations until event \mathcal{E} is satisfied. Under the assumption that no individual reaches more than $(\sigma + \delta)n$ leading 1-bits, the bits after position $(\sigma + \delta)n$ will be uniformly distributed. Hence, by a Chernoff bound, the probability that a given individual has more than $2\delta n/3$ 1-bits in the interval from $n(\sigma + \delta)$ to $n(\sigma + 2\delta)$ is exponentially small. By Theorem 3 and Theorem 4, the expectation is $\mathbf{E}[\tau_c] = O(n^2)$. To find the optimum while event \mathcal{E} is satisfied, it suffices to select an individual with rank between 0 and γ , and flip the leading $k + 3$ 1-bits, an event which happens in each generation with probability at least

$$\begin{aligned} 1 - \left(1 - \frac{\beta(\gamma)}{n^{k+3}}\right)^\lambda &\geq 1 - \exp\left(-\frac{\lambda\beta(\gamma)}{n^{k+3}}\right) \\ &\geq 1 - \frac{1}{1 + \frac{\lambda\beta(\gamma)}{n^{k+3}}} \\ &\geq \frac{\lambda\beta(\gamma)}{n^{k+3} + \lambda\beta(\gamma)} \\ &\geq \frac{\lambda}{2n^{k+3}}. \end{aligned}$$

By Theorem 2, with probability $e^{-\Omega(\lambda)}$, the γ -ranked individual has either less than $(\sigma - \delta)n$ leading 1-bits or the 0-ranked individual has more than $(\sigma + \delta)n$ leading 1-bits in the following generation. Hence, the expected number of generations τ conditional on event \mathcal{E} is

$$\begin{aligned} \mathbf{E}[\tau \mid \mathcal{E}] &\leq \left(1 - \frac{\lambda}{2n^{k+3}} - e^{-\Omega(\lambda)}\right) \cdot (1 + \mathbf{E}[\tau \mid \mathcal{E}]) \\ &\quad + e^{-\Omega(\lambda)} \cdot (\mathbf{E}[\tau_c] + \mathbf{E}[\tau \mid \mathcal{E}]) \\ &\leq (1/\lambda) \cdot (2n^{k+3} + 2n^{k+5} \cdot e^{-\Omega(\lambda)}) \\ &= O(n^{k+3}/\lambda). \end{aligned}$$

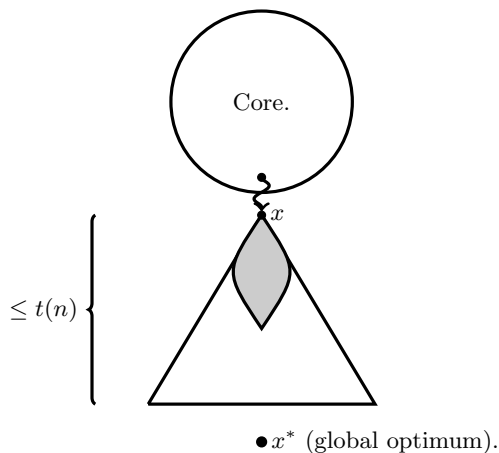


Figure 4: Non-selective family tree (triangle) of the family tree (gray) rooted in individual x .

The unconditional runtime is therefore

$$\mathbf{E}[T] = \lambda \cdot (\mathbf{E}[\tau_c] + \mathbf{E}[\tau \mid \mathcal{E}]) = O(n^{k+3}).$$

□

Theorem 5 is conditional on the event that no individual in the population reaches more than $(\sigma + \delta)n$ leading 1-bits. The probability of this event is unknown, but based on the results in Section 4.1.1, one can conjecture that the probability of this event is exponentially close to one. In particular, Theorem 2 implies that for any arbitrarily small constant $\gamma > 0$, none of the individuals ranked between γ and 1 will reach more than $(\sigma + \delta)n$ leading 1-bits. Furthermore, if Conjecture 6 in Section 4.5 holds, then the extra condition in Theorem 5 can be removed.

4.3 Non-selective Family Trees

Our next goal is to prove that there is an exponentially small probability that any individual reaches far beyond the equilibrium position within exponential time. However, Theorems 2 and 3 assume that the rank parameter γ is a constant, and cannot be used to analyse the behaviour of single “stray” individuals, including the position of the fittest individual (i.e. $\gamma = 0$). This is because the tail inequalities obtained by the Chernoff bounds used in the proofs of these theorems are too weak when $\gamma = o(1)$.

To analyse stray individuals, we will apply the notion of *family trees* as described by Witt [17], although in a slightly different way. A family tree has as its root a given individual x in some generation t , and the nodes in each level k correspond to the subset of the population in generation $t + k$ defined in the following way. An individual y in generation $t + k$ is a member of the family tree if and only if it was generated by selection and mutation of an individual z that belongs to level $t + k - 1$ of the family tree. In this case, individual z is the parent node of individual y . If there is a path from an individual z at level k to an individual y at level $k' > k$, then individual y is said to be a *descendant* of individual z , and individual z is an *ancestor* of individual y . A path in the family tree is called a *lineage*. A family tree is said to become *extinct* in generation $t + t(n) + 1$ if none of the individuals in level $t(n)$ of the tree were selected. In this case, $t(n)$ is called the *extinction time* of the family tree.

The idea for proving that stray individuals do not reach a given part of the search space can be described informally using Figure 4. One defines a certain subset of the search space called the *core* within which the majority of the population is confined with overwhelming probability. In our case, an appropriate core can be defined using Theorems 2 and 3. One then focuses on the family trees that are outside this core, but where the root has a parent in the core. Note that some descendants of the root may re-enter the core. We therefore prune the family tree to those descendants which are always outside the core. More formally, the pruned family tree contains node x if and only if x belongs to the original family tree, and x and all its ancestors are outside the core.

We would then like to analyse the positions of the individuals that belong to the pruned family tree. However, it is non-trivial to calculate the exact shape of this family tree. Let random variable O_x denote the number of offspring of individual x . Clearly, the distribution of O_x depends on how x is ranked within the population. Hence, different parts of the pruned family tree may grow at different rates, which can influence the position and shape of the family tree. To simplify the analysis, we embed the pruned family tree into a larger family tree which we call the *non-selective family tree*. This family tree has the same root as the real pruned family tree, however it grows through a modified selection process. In the real pruned family tree, the individuals have different numbers of offspring according to their rank in the population. In the non-selective family tree, the offspring distribution O_x of all individuals x is identical to the offspring distribution O_z of an individual z which is best ranked among individuals outside the core. Hence, each individual in the non-selective family tree has at least as many offspring as in the real family tree. The real family tree will therefore occur as a sub-tree in the non-selective family tree. Furthermore, the probability that the real family tree reaches a given part of the search space, is upper bounded by the probability that the non-selective family tree reaches this part of the search space. A related approach, where faster growing family trees are analysed, is described by Jägersküpfer and Witt [11].

Approximating the family tree by the non-selective family tree has three important consequences. The *first* consequence is that the non-selective family tree can grow faster than the real family tree, and in general beyond the population size λ of the original process. The *second* consequence is that since all individuals in the non-selective family tree have the same offspring distribution, no individual in the non-selective family tree has any selective advantage, hence the name non-selective family tree. The behaviour of the family tree is therefore independent of the fitness function, and each lineage fluctuates randomly in the search space according to the bits flipped by the mutation operator. Such mutation random walks are easier to analyse than the real search process. To bound the probability that such a mutation random walk enters a certain area of the search space, it is necessary to bound the extinction time $t(n)$ of the non-selective family tree. The *third* consequence is that the sequence of random variables $Z_{t \geq 0}$ describing the number of elements in level t of the non-selective family tree is a discrete time branching process [8]. We can therefore apply the techniques that have been developed to study such processes to bound $t(n)$.

Definition 3 (BRANCHING PROCESS [8]). *A single type branching process is a Markov process Z_0, Z_1, \dots which for all $n \geq 0$, is given by $Z_{n+1} := \sum_{i=1}^{Z_n} \xi_i$, where $\xi_i \in \mathbb{N}_0$ are i.i.d. random variables having $\mathbf{E}[\xi] =: \rho$.*

A branching process can be thought of as a population of identical individuals, where each individual survives exactly one generation. Each individual produces ξ offspring independently of the rest of the population during its lifetime, where ξ is a random variable with expectation ρ . The random variable Z_t denotes the population size in generation t . Clearly, if $Z_t = 0$ for some t , then $Z_{t'} = 0$ for all $t' \geq t$. The following lemma gives a simple bound on the size of the population after $t \geq 1$ generations.

Lemma 2. *If Z_0, Z_1, \dots is a single-type branching process with $Z_0 := 1$ and mean number of offspring per individual ρ , then $\Pr[Z_t \geq k] \leq \rho^t/k$ for any $k > 0$.*

PROOF. Markov's inequality gives

$$\begin{aligned} \Pr[Z_t \geq k] &\leq \mathbf{E}[Z_t]/k \\ &= \mathbf{E}[\mathbf{E}[Z_t | Z_{t-1}]]/k \\ &= \rho/k \cdot \mathbf{E}[Z_{t-1}] \\ &= \rho^t/k \cdot \mathbf{E}[Z_0]. \end{aligned}$$

□

Clearly, the expected number of offspring ρ is important for the fate of a branching process. For $\rho < 1$, the process is called *sub-critical*, for $\rho = 1$, the process is called *critical*, and for $\rho > 1$, the process is called *super-critical*.

4.4 Too high Selection Pressure

In this section, it is proved that $\text{SELPRES}_{\sigma,k}$ is hard for Linear Ranking EA when the ratio between parameters η and χ is sufficiently large. The proof idea is to show that in the equilibrium position, a majority of the individuals have considerably more than $(\sigma + \delta)n$ leading 1-bits. Individuals close to the optimum are therefore less likely to be selected. First, it is shown in Propositions 1, 2 and 3 that there is a non-negligible probability that the equilibrium position is reached before the optimum is found. In the following, we will call any search point with prefix x an x -individual.

Proposition 1. *If the initial generation contains more than $2^{-k-4} \cdot \lambda 1^{k+3}$ -individuals, then the expected number of generations until the 1^{k+3} -individuals occupy more than half of the population is $O(1)$.*

PROOF. The 1^{k+3} -individuals are fitter than any other non-optimal individuals. Hence, if the fraction of 1^{k+3} -individuals in the population is $2^{-k-4} < \gamma \leq 2^{-1}$, then the expected fraction of 1^{k+3} -individuals in the following generation is at least $r\gamma$, where

$$r \geq \frac{\beta(\gamma)}{\gamma} \cdot \left(1 - \frac{\chi}{n}\right)^{k+3} \geq \frac{\eta+1}{2} \cdot \left(1 - \frac{\chi}{n}\right)^{k+3}.$$

Hence, for sufficiently large n , there exists a constant $c > 0$ such that $r > 1 + c$. Starting with a fraction of $\gamma > 2^{-k-4}$ 1^{k+3} -individuals, as long as the fraction of 1^{k+3} -individuals is below $1/2$, the expected ratio of 1^{k+3} -individuals in generation t is at least $\gamma \cdot (1+c)^t$. Hence, the expected number of generations t until the 1^{k+3} -individuals occupy at least half of the population satisfies $2^{-k-4} \cdot (1+c)^t \geq 1/2$, which holds for $t = O(1)$. □

Proposition 2. *If the Linear Ranking EA with population size $\lambda \leq n^k$ is applied to $\text{SELPRES}_{\sigma,k}$, then the probability that the first individual that finds the optimum has a 1^{k+3} -individual as ancestor is $1 - e^{-\Omega(n)}$.*

PROOF. We apply the idea of non-selective family trees. Define the core as the set of 1^{k+3} -individuals. We will now bound the probability that any given non-selective family tree outside the core finds the optimum.

By Chernoff bounds, there is an exponentially large probability that the initial generation contains at least $\lambda \cdot 2^{-k-4}$ 1^{k+3} -individuals. Hence, By Proposition 1, after a constant number of generations, any individual outside the core has rank higher than $1/2$. The expected number of times an individual with rank γ is selected during one generation is no more than $\alpha(\gamma)$ as given in Eq. (1). Hence for selection pressure $\eta > 1$, the expected number of times an individual with rank higher than $1/2$ is selected is less than ρ for some constant $\rho < 1$. For a given family tree, let random variable X_t denote the number of individuals in the non-selective family tree in generation t , where $X_0 = 1$ corresponds to the single root and assume that every family member has exactly ρ expected offspring. Then X_t is a single type branching process [8], and the expected number of family members in generation t can be bounded by $\mathbf{E}[X_t] = \mathbf{E}[\mathbf{E}[X_t | X_{t-1}]] \leq \rho \cdot \mathbf{E}[X_{t-1}] \leq \rho^t$.

We will now bound the number of different lineages that exist within the at most λ family trees. Note that the number of different lineages within one family tree equals the number of leaf nodes in the family tree, which is trivially bounded by the product of the height H of the tree and the maximal width W of the family tree. The family tree height is the extinction time of the family tree, and the family tree width corresponds to the maximum number of alive family members within one generation. The probability that the height is at most n can be bounded using Markov's inequality to $\Pr[H \leq n] = 1 - \Pr[X_n \geq 1] = 1 - \mathbf{E}[X_n] = 1 - \rho^n$. Furthermore, the probability that the width exceeds $e^{c_w n}$ can be bounded by $\Pr[W \leq e^{c_w n}] = \Pr[\max_t X_t \leq e^{c_w n}] \geq 1 - e^{-c_w n}$, where $c_w > 0$ is any constant. Hence, by union bound, the probability that the number of lineages in the at most λ family trees is less than $\lambda n e^{c_w n}$ is at least $(1 - \lambda \cdot e^{-c_w n}) \cdot (1 - \lambda \cdot \rho^n) = 1 - e^{-\Omega(n)}$.

We will now bound the probability that one given lineage outside the core finds the optimum, conditional on the event that the lineage survives at most n generations. The root of the family tree corresponds to an individual in the first generation of the EA, which is a bitstring sampled uniformly at random. Hence, by a Chernoff bound, with probability $1 - e^{-\Omega(n)}$, the number of 0-bits in the interval from $k+4$ to $(\sigma - \delta)n$ is at least $(\sigma - \delta)n/3$. In order to reach the optimum, it is necessary that all these 0-bits are flipped into 1-bits. However, the probability that a given of these bits has not been flipped within n generations is $(1 - \chi/n)^n > c'$ for some constant $c' > 0$. Hence, the probability that all of the at least $(\sigma - \delta)n/3$ 0-bits have been flipped within n generations is less than $(1 - c')^{(\sigma - \delta)n/3} \leq e^{-cn}$ for some constant $c > 0$.

Finally, the probability that any of the $\lambda n e^{c_w n}$ lineages finds the optimum is by union bound at most $\lambda n e^{c_w n} e^{-cn} = e^{-\Omega(n)}$ for sufficiently small c_w . Hence, the unconditional probability that none of the lineages finds the optimum is $1 - e^{-\Omega(n)}$. □

Proposition 3. For any constant $r > 0$, the probability that the linear ranking EA with population size $\lambda \leq n^k$ has not found the optimum of SELPRES $_{\sigma,k}$ within rn^2 generations is $\Omega(1)$.

PROOF. By Proposition 2, with exponentially high probability, the first globally optimal individual has an 1^{k+3} -individual as ancestor. We therefore bound the probability of finding the optimum, by the probability that an 1^{k+3} -individual has a 0^{k+3} -individual as descendant within the first rn^2 generations, or equivalently within $\lambda rn^2 \leq rn^{k+2}$ selection steps. We distinguish between the following two cases.

Case 1: The 0^{k+3} -individual is created directly from a 1^{k+3} -individual by mutating the $k+3$ leading 1-bits simultaneously. The probability that this event happens in any mutation step is no more than $(\chi/n)^{k+3}$, and the probability that this does not happen in rn^{k+2} selection steps is by union bound $1 - O(1/n)$.

Case 2: The 0^{k+3} -individual is created by first mutating a 1^{k+3} -individual into an intermediary individual y that has $m, 1 \leq m < k$ 1-bits among the first k bits, and then individual y has an 0^{k+3} -individual as descendant.

To analyse this situation, we apply the method of non-selective family trees where the core is defined as the set of 1^{k+3} -individuals. We consider a family tree that is rooted in y . Let random variable F denote the number of such family trees occurring in rn^{k+2} selection steps. The probability of creating a new family tree from a core member in one selection step is $\binom{k}{k-m} \cdot (\chi/n)^{3+k-m}$, hence the expected value of F is no more than $(c/2) \cdot n^{m-1}$ where $c := 2 \cdot (2\chi)^k$. The probability of the event \mathcal{F} that F is less than cn^{m-1} is by Markov's inequality bounded by $\Pr[\mathcal{F}] = 1 - \Pr[F \geq cn^{m-1}] \geq 1/2$.

For a given family tree, let the random variable X_t denote the number of family members in generation t of the lifetime of the family tree, where $X_0 = 1$. Following the ideas in Proposition 2 for family trees outside the core, the expected number of offspring per family member is bounded by a constant $\rho < 1$, and one obtains $\mathbf{E}[X_t] \leq \rho^t$, and $\mathbf{E}[\max_t X_t] \leq 1$. The extinction time D of any given such family tree can now be bounded by

$$\begin{aligned} \Pr[D \leq m \ln n] &\geq 1 - \Pr[X_{m \ln n} \geq 1] \\ &\geq 1 - \mathbf{E}[X_{m \ln n}] \\ &\geq 1 - \rho^{m \ln n} \\ &= 1 - O(n^{-m}). \end{aligned}$$

And the probability of event \mathcal{D} , that all family trees are extinct within $m \ln n$ generations, is bounded by $\Pr[\mathcal{D} | \mathcal{F}] \geq (1 - O(n^{-m}))^{cn^{m-1}} > 1/e$.

Let random variable P denote the number of paths from root to leaf within the forest of all family trees that arise within rn^2 generations. Conditional on the events \mathcal{D} and \mathcal{F} , the random value P is bounded by $P \leq m \ln n \sum_{i=1}^{cn^{m-1}} W_i$, where random variable W_i denotes the maximal width (ie. maximum number of living family members during a generation) of family tree i . By Markov's inequality, the probability of the event \mathcal{P} that P is less than $2mcn^{m-1} \ln n$ is

bounded by

$$\begin{aligned} \Pr[\mathcal{P} | \mathcal{F}, \mathcal{D}] &\geq 1 - \Pr[P \geq 2mcn^{m-1} \ln n | \mathcal{F}, \mathcal{D}] \\ &\geq 1 - \frac{\mathbf{E}[P | \mathcal{F}, \mathcal{D}]}{2mcn^{m-1} \ln n} \\ &\geq 1 - \frac{(m \ln n) \cdot (cn^{m-1}) \cdot \mathbf{E}[\max_t X_t]}{2mcn^{m-1} \ln n} \\ &\geq 1/2. \end{aligned}$$

We now calculate the probability that a given path of length at most $m \ln n$ finds the optimum. The probability of flipping a given bit within $m \ln n$ mutation steps is by union bound less than $\chi m^2 \ln n/n$, and the probability that all the m remaining 1-bits have been flipped is by union bound less than $(\chi m^2 \ln n/n)^m$.

The probability that any of the at most $2mcn^{m-1} \ln n$ paths finds the optimum, conditional on the events \mathcal{F} and \mathcal{D} is by union bound less than $(2mcn^{m-1} \ln n)(\chi m^2 \ln n/n)^m = O((\ln n)^{m+1}/n)$. Hence, the unconditional probability that the optimum has not been found within the first rn^2 generations is $\Omega(1)$. \square

Definition 4. Let T be any family tree and ξ any constant $\xi, 0 < \xi < 1$. The ξn_{\leq} -pruning of family tree T is the family tree consisting of any member x of T such that x and all the ancestors of x in T have at most ξn leading 1-bits.

Proposition 4. Let $\eta > 2 \exp(\chi(\sigma + 3\delta)) - 1 + \delta$, and x any individual which has less than ξn leading 1-bits, with $\xi := \sigma + 2\delta$. If the $(1 + \delta)/2$ -ranked individual has at least $(\xi + \delta)n$ leading 1-bits, then the probability that the ξn_{\leq} -pruned family tree of individual x is extinct in generation $t_0 + n$ is exponentially high $1 - e^{-\Omega(n)}$.

PROOF. Let t_0 denote the generation number when the family tree rooted in individual x occurs, and let random variable X_t denote the number of members of the pruned family tree in generation $t_0 + t$, where the initial family size is $X_0 := 1$. For $\gamma := (1 + \delta)/2$, one has $\ln(\beta(\gamma)/\gamma)/\chi \geq \sigma + 3\delta = \xi + \delta$, hence by Theorem 2, with exponentially high probability, the individuals ranked $(1 + \delta)/2$ or better will have more than ξn leading 1-bits. Therefore, the members of the ξn_{\leq} -pruned family tree must have ranks at least $(1 + \delta)/2$. The expected number of offspring for a given member of the pruned family tree in generation $t_0 + t$ is therefore no more than $\alpha((1 + \delta)/2)$.

By Lemma 2, the probability that the pruned family tree is not extinct after n generations is $\Pr[X_n \geq 1] \leq \alpha((1 + \delta)/2)^n = (1 - (\eta - 1) \cdot \delta)^n = e^{-\Omega(n)}$. \square

Proposition 5. Let $\eta > 2 \exp(\chi(\sigma + 3\delta)) - 1 + \delta$ and $\lambda = \Omega(n)$. If the $(1 + \delta)/2$ -ranked individual reaches at least $(\sigma + 3\delta)n$ leading 1-bits before the optimum has been found, then the probability that the optimum is found within e^{cn} generations is exponentially small $e^{-\Omega(n)}$, where c is a constant.

PROOF. Define $\xi := \sigma + 2\delta$, and define the *core set*, as the set of search points with more than ξn leading 1-bits. By Theorem 2, the probability of the event that the $(1 + \delta)/2$ -ranked individual has less than $\sigma + 2\delta$ leading 1-bits in the next generation is $e^{-\Omega(n)}$, and by union bound, the probability that this happens within e^{cn} generations is $e^{-\Omega(n)}$ for sufficiently small c . In the following, we therefore assume that this event does not happen within e^{cn} steps.

The event where an individual x belonging to the core set has offspring with less than ξn leading one-bits is called a *trial*. A trial is called *successful* if any member of the ξn_{\leq} -pruned family tree from individual x is optimal. We will now define a non-selective family tree, and bound the success probability of a trial by the probability of finding a global optimum within the family tree.

We first bound the number of different lineages in the family tree. Analogously to the proof of Proposition 2, the number of different lineages in an family tree can be bounded by the product of the height H and the width W of the non-selective family tree. By Proposition 4, with exponentially high probability, the ξn_{\leq} -pruned family tree is extinct within n generations, hence the height is bounded by n . To bound the width, note that each individual in the non-selective family tree has rank at least $(1 + \delta)/2$. Hence, following the proof of Proposition 4, each individual receives in expectation less than $\alpha((1 + \delta)/2)$ number of offspring per generation. Denoting the number of individuals in the non-selective family tree in generation t with X_t , the expected size of the non-selective family tree in generation t can be bounded by $\mathbf{E}[X_t] \leq (1 - \delta)^t < 1$, using the same calculation as in the proof of Proposition 4. The probability that the non-selective family tree grows beyond $(1/p)^{(\ln \sigma/6)n}$ individuals within n generations, where $0 < p < 1$ is a constant that will be specified later, can now be bounded using Markov's inequality by

$$\begin{aligned} \Pr \left[W \geq (1/p)^{(\ln \sigma/6)n} \right] &= \\ \Pr \left[\max_{1 \leq t \leq n} X_t \geq (1/p)^{(\ln \sigma/6)n} \right] & \\ \leq p^{(\ln \sigma/6)n} \cdot \mathbf{E} \left[\max_{1 \leq t \leq n} X_t \right] &\leq e^{-c'n}, \end{aligned}$$

where $c' := (\ln \sigma/6) \ln(1/p)$. Any trial that has more than $(1/p)^{(\ln \sigma/6)n}$ different random walks from x to the leaf nodes, is called *successful*. Consider any other trial. By the definition of a trial, the initial individual x has only 1-bits in the interval between $(\sigma + \delta) \cdot n$ and ξn . In order to reach the global optimum, it is necessary that the random walk flips at least $\delta n/3$ 1-bits in this interval, without flipping any 1-bits before index $(\sigma - \delta) \cdot n$. Instead of considering the bit-flips that occur during one generation, we note the positions of all the bit-flips that occur during n generations, and ignore the generation number in which the bit-flips occurred. Clearly, in order to obtain the optimum, at least $\delta n/3$ bit-flips must have occurred. However, the position of each bit-flip is uniform from 1 to n , and the probability that a given bit-flip occurred before position $(\sigma - \delta) \cdot n$, is $(\sigma - \delta)$. Hence, the probability that none of the $\delta n/3$ bit-flips occurs before position $(\sigma - \delta)n$ is less than $p^{(\ln \sigma/3) \cdot n}$, where p is defined to be $p := 1 - \sigma + \delta$. The number of lineages in one trial is bounded by $W \cdot H \leq (1/p)^{(\ln \sigma/6)n} n$. By union bound, the probability that any of the random walks in the trial finds the global optimum is bounded from above by

$$n \cdot (1/p)^{(\ln \sigma/6)n} \cdot p^{(\ln \sigma/3) \cdot n} = e^{-c'n + \ln n}.$$

The probability that at least one of the at most $\lambda \cdot e^{cn}$ trials is successful, is no more than

$$\lambda \cdot e^{cn} \cdot e^{-c'n + \ln n} = e^{-\Omega(n)}$$

when c is sufficiently small. \square

Theorem 6. *If $\eta > 2 \exp(\chi(\sigma + 3\delta + \epsilon)) - 1 + \delta$, where $\epsilon > 0$ is any constant, and $n \leq \lambda \leq n^k$, then the expected runtime of the Linear Ranking EA on $\text{SELPRES}_{\sigma,k}$ is $e^{\Omega(n)}$.*

PROOF. By Theorem 3 and Markov's inequality, there is a constant probability that the $\gamma := (1 + \delta)/2$ -ranked individual has reached at least

$$\begin{aligned} n(\ln(\beta(\gamma)/\gamma)/\chi - \epsilon) &\geq n(\ln((\eta + 1 - \delta)/2)/\chi - \epsilon) \\ &\geq n(\sigma + 3\delta) := \xi n \end{aligned}$$

leading 1-bits within rn^2 generations, for some constant r . By Proposition 3, the probability that the optimum has not been found within the first rn^2 generations is $\Omega(1)$. If the optimum has not been found before the $(1 + \delta)/2$ -ranked individual has ξn leading 1-bits, then by Proposition 5, the expected runtime is $e^{\Omega(n)}$. The unconditional expected runtime of the linear ranking EA is therefore $e^{\Omega(n)}$. \square

4.5 Too low Selection Pressure

It is natural to ask whether there exists an analogue to Theorem 6 for parameter settings where the equilibrium position $n(\ln \eta)/\chi$ is significantly below $(\sigma - \delta)n$. I.e., is $\text{SELPRES}_{\sigma,k}$ also hard when the selection pressure is too low? We conjecture that this is the case.

Conjecture 1. *If $\eta = \exp(\chi(\sigma - \delta)) - \epsilon$ for any $\epsilon > 0$, then the probability that Linear Ranking EA with population size $\lambda < n^k$ will find the optimum of $\text{SELPRES}_{\sigma,k}$ within e^{cn} generations is $e^{-\Omega(n)}$, where c is a constant.*

Unfortunately, modelling family trees outside the core using a single-type branching process as in the previous section will not work in this case. E.g., a direct application of the ideas above would be to define the core as the set of search points with less than $n \ln(\eta\kappa)/\chi$ leading 1-bits for some small constant $\kappa > 1$. The problem with this approach is that the number of leading 1-bits can potentially increase significantly by flipping a single 0-bit. A more detailed characterisation of the process may therefore be needed to analyse this case.

5. CONCLUSION

The objective of this paper has been to better understand the relationship between mutation and selection in EAs, and in particular to what degree this relationship can have an impact on the runtime. To this end, we have rigorously analysed the runtime of a non-elitistic population-based evolutionary algorithm that uses linear ranking selection and bitwise mutation on a family of fitness functions. We have focused on the effects of two parameters of the EA, η which controls the selection pressure, and χ which controls the bitwise mutation rate χ/n .

The theoretical results show that there exist fitness functions where the parameter settings of selection pressure η and mutation rate χ have a dramatic impact on the runtime. A small change in the mutation rate or the selection pressure can increase the runtime of the EA from a small polynomial (ie highly efficient), to exponential (ie. highly inefficient). The results show that the EA will have exponential runtime if the selection pressure becomes too high, or the mutation rate becomes too low. The polynomial upper bound is conditional on an assumption about the maximum number of leading 1-bits obtained by the best individual in

the population. If this assumption holds, then to achieve polynomial runtime on the problem, the settings of these parameters must be adjusted to a narrow region of the parameter space, as illustrated in Figure 2. The critical factor which determines whether the EA is efficient on the problem is not the individual parameter settings of χ or η , but rather the ratio $\ln(\eta)/\chi$ between these two parameters. Hence, a too high mutation rate χ can be balanced by increasing the selection pressure η , and a too low selection pressure η can be balanced by decreasing the mutation rate χ . The analysis also shows that the position of the critical region in parameter space in which the EA is efficient is problem dependent. Hence, the EA may be efficient with a given mutation rate and selection pressure on one problem, but be highly inefficient with the same parameter settings on another problem. There is therefore no optimal balance between the selection pressure and the mutation rate which is robust on all problems.

Informally, these results can be explained due to the occurrence of an equilibrium state into which the non-elitistic population enters after a certain time. In this state, the EA makes no further progress, even though there is a fitness gradient in the search space. The position in the search space in which the equilibrium state occurs depends on the mutation rate and the selection pressure. When the number of new good individuals added to the population by selection equals the number of good individuals destroyed by mutation, then the population makes no further progress. If the equilibrium state occurs close to the global optimum, then the EA is efficient. If the equilibrium state occurs far from the global optimum, then the EA is inefficient.

The results are theoretically significant because the impact of selection pressure on the runtime of EAs has not previously been analysed. Furthermore, there exists few results on population-based EAs, in particular those that employ both a parent and an offspring population. In addition, the runtime analysis applied techniques that are new to the field. In particular, the behaviour of the main part of the population and stray individuals are analysed separately. The analysis of stray individuals is achieved using a concept which we call non-selective family trees, which are then analysed as single-type branching processes. These new techniques may potentially be applicable to a wider set of EAs and fitness functions. Finally, our analysis answers a challenge by Happ et al. [9], to analyse a population-based EA using a non-elitistic selection mechanism. The results also shed some light on the possible reasons for the difficulty of parameter tuning in practical applications of evolutionary algorithms. The optimal parameter settings can be problem dependent, and very small changes in the parameter settings can have big impacts on the efficiency of the algorithm.

A challenge for future experimental work is to design and analyse strategies for dynamically adjusting the mutation rate and selection pressure. Can self-adaptive EAs be robust on problems like those that are described in this paper? Future theoretical work should try to prove Conjecture 1. Furthermore, it would be interesting to extend the analysis to other selection mechanisms than linear ranking selection, and to EAs that apply a crossover operator.

6. ACKNOWLEDGEMENTS

The extraordinary detailed comments provided by one of the anonymous reviewers helped to improve the paper signif-

icantly. The authors would also like to thank Tianshi Chen for discussions about selection mechanisms in evolutionary algorithms and Roy Thomas, Lily Kolotilina and Jon Rowe for discussions related to branching processes.

This work was supported by the EPSRC under grant no. EP/D052785/1.

7. REFERENCES

- [1] Thomas Bäck. Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the 1st IEEE Conf. on Evolutionary Computation*, pages 57–62. IEEE Press, 1994.
- [2] Tobias Blickle and Lothar Thiele. A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, 4(4):361–394, 1996.
- [3] Erick Cantu-Paz. Order statistics and selection methods of evolutionary algorithms. *Information Processing Letters*, 82(1):15–22, 2002.
- [4] Tianshi Chen, Jun He, Guangzhong Sun, Guoliang Chen, and Xin Yao. A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2009. To appear.
- [5] Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the (1+1) Evolutionary Algorithm. *Theoretical Computer Science*, 276:51–81, 2002.
- [6] Agoston E. Eiben and C. A. Schippers. On evolutionary exploration and exploitation. *Fundam. Inf.*, 35(1-4):35–50, 1998.
- [7] David E. Goldberg and Kalyanmoy Deb. A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kaufmann, 1991.
- [8] Patsy Haccou, Peter Jagers, and Vladimir Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, 2005.
- [9] Edda Happ, Daniel Johannsen, Christian Klein, and Frank Neumann. Rigorous analyses of fitness-proportional selection for optimizing linear functions. In *GECCO '08: Proceedings of the 10th annual conference on Genetic and evolutionary computation*, pages 953–960, New York, NY, USA, 2008. ACM.
- [10] Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.
- [11] Jens Jägersküpfer and Carsten Witt. Rigorous runtime analysis of a $(\mu+1)$ ES for the sphere function. In *GECCO '05: Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 849–856, New York, NY, USA, 2005. ACM.
- [12] Tatsuya Motoki. Calculating the expected loss of diversity of selection schemes. *Evolutionary Computation*, 10(4):397–422, 2002.
- [13] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

- [14] Ruhul Sarker, Masoud Mohammadian, and Xin Yao, editors. *Evolutionary Optimization*. Kluwer Academic Publishers, 2002.
- [15] Dirk Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm. *Evolutionary Computation*, 1:25–49, 1993.
- [16] Darrell Whitley. The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In J. D. Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, San Mateo, CA, 1989. Morgan Kaufman.
- [17] Carsten Witt. Runtime Analysis of the $(\mu + 1)$ EA on Simple Pseudo-Boolean Functions. *Evolutionary Computation*, 14(1):65–86, 2006.
- [18] Carsten Witt. Population size versus runtime of a simple evolutionary algorithm. *Theoretical Computer Science*, 403(1):104–120, 2008.