On Proportions of Fit Individuals in Population of Mutation-Based Evolutionary Algorithm with Tournament Selection

Anton V. Eremeev[*]   eremeev@ofim.oscsbras.ru
Sobolev Institute of Mathematics, Omsk Branch, 13 Pevtsov str.,
Omsk, 644099, Russia, and Omsk State University n.a. F.M. Dostoevsky,
Mira avenue, 55A, Omsk, 630077, Russia

Abstract

In this article, we consider a fitness-level model of a non-elitist mutation-only evolutionary algorithm (EA) with tournament selection. The model provides upper and lower bounds for the expected proportion of the individuals with fitness above given thresholds. In the case of so-called monotone mutation, the obtained bounds imply that increasing the tournament size improves the EA performance. As corollaries, we obtain an exponentially vanishing tail bound for the Randomized Local Search on unimodal functions and polynomial upper bounds on the runtime of EAs on the 2-SAT problem and on a family of Set Cover problems proposed by E. Balas.

1 Introduction

Evolutionary algorithms are randomized heuristic algorithms employing a population of tentative solutions (individuals) and simulating an evolutionary type of search for optimal or near-optimal solutions by means of selection, crossover, and mutation operators. The evolutionary algorithms with crossover operator are usually called genetic algorithms (GAs). Evolutionary algorithms in general have a more flexible outline and include genetic programming, evolution strategies, estimation of distribution algorithms, and other evolution-inspired paradigms. Evolutionary algorithms are now frequently used in areas of operations research, engineering, and artificial intelligence.

Two major outlines of an evolutionary algorithm are the *elitist* evolutionary algorithm, that keeps a certain number of most promising individuals from the previous iteration, and the *non-elitist* evolutionary algorithm, that computes all individuals of a new population independently using the same randomized procedure. In this article, we focus on the non-elitist case.

One of the first theoretical results in the analysis of non-elitist GAs is the Schemata Theorem (Goldberg, 1989) which gives a lower bound on the expected number of individuals from some subsets of the search space (schemata) in the next generation, given

---

the current population. A significant progress in understanding the dynamics of GAs with non-elitist outline was made in Vose (1995) by means of dynamical systems. However most of the findings in Vose (1995) apply to the infinite population case, and it is not clear how these results can be used to estimate the applicability of GAs to practical optimization problems. A theoretical possibility of constructing GAs that provably optimize an objective function with high probability in polynomial time was shown in Vitányi (2000) using rapidly mixing Markov chains. However, Vitányi (2000) provides only a very simple artificial example where this approach is applicable and further developments in this direction are not known to us.

One of the standard approaches to studying evolutionary algorithms in general is based on the *fitness levels* (Wegener, 2002). In this approach, the solution space is partitioned into disjoint subsets, called fitness-levels, according to values of the fitness function. In Lehre (2011), the fitness-level approach was first applied to upper-bound the runtime of non-elitist mutation-only evolutionary algorithms. Here and below, by the runtime we mean the expected number of fitness evaluations made until an optimum is found for the first time. Upper bounds of the runtime of non-elitist GAs, involving the crossover operators, were obtained later in Corus et al. (2014) and Eremeev (2017). The runtime bounds presented in Corus et al. (2014) and Lehre (2011) are based on the drift analysis. In Moraglio and Sudholt (2015), a runtime result is proposed for a class of convex search algorithms, including some non-elitist crossover-based GAs without mutation, on the so-called concave fitness landscapes.

In this article, we consider the non-elitist evolutionary algorithm, which uses a tournament selection and a mutation operator but no crossover. The s-tournament selection randomly chooses s individuals from the existing population and selects the best one of them (see, e.g., Thierens and Goldberg, 1994). The mutation operator is viewed as a randomized procedure, which computes one offspring with a probability distribution depending on the given parent individual. In this article, evolutionary algorithms with such outline are denoted as EA. We study the probability distribution of the EA population with regards to a set of fitness levels. The estimates of the EA behavior are based on a priori known parameters of a mutation operator. Using the proposed model we obtain upper and lower bounds on expected proportion of the individuals with fitness above certain thresholds. The lower bounds are formulated in terms of linear algebra and resemble the bound in the Schemata Theorem (Goldberg, 1989). Instead of schemata, here we consider the sets of genotypes with the fitness bounded from below. Besides that, the bounds obtained in this article may be applied recursively up to any given iteration.

The lower bounds on expected proportions of sufficiently fit individuals at iteration t also imply the lower bounds on probabilities of finding a genotype with fitness above a specified threshold at any given iteration t. Such results are closely related to the area of fixed budget computations, where one has a fixed budget of fitness evaluations that may be spent and the question is how good a solution one can expect to find with this budget (Jansen and Zarges, 2012).

This article pays particular attention to a special case when mutation is *monotone*. Informally speaking, a mutation operator is monotone if fitter parents have a higher probability of producing fit offspring. One of the most well-known examples of monotone mutation is the bitwise mutation in the case of OneMax fitness function. As shown in Borisovsky and Eremeev (2008), in the case of monotone mutation, one of the most simple evolutionary algorithms, known as the $(1+1)$ EA has the best-possible performance in terms of runtime and probability of finding the optimum.

In the case of monotone mutation, the lower bounds on expected proportions of the individuals turn into equalities for the trivial evolutionary algorithm (1,1) EA. This

implies that the tournament selection at least has no negative effect on the EA performance in such a case. This observation is complemented by the asymptotic analysis of the EA with monotone mutation indicating that, given a sufficiently large population size and some technical conditions, increasing the tournament size s, always improves the EA performance.

As corollaries of the general lower bounds on expected proportions of sufficiently fit individuals, we obtain polynomial upper bounds on the Randomized Local Search (RLS) runtime on unimodal functions and upper bounds on the runtime of EAs on 2-SAT problem and on a family of Set Cover problems proposed by Balas (1984). Unlike the upper bounds on the runtime of evolutionary algorithms with tournament selection from Corus et al. (2014), Eremeev (2017) and Lehre (2011), which require sufficiently large tournament size, the upper bounds on runtime obtained here hold for any tournament size.

The rest of the article is organized as follows. In Section 2, we give a formal description of the considered EA, introduce an approximating model of the EA population, and define some required parameters of the probability distribution of a mutation operator in terms of fitness levels. In Section 3, using the model from Section 2, we obtain lower and upper bounds on expected proportions of genotypes with fitness above some given thresholds. Section 4 is devoted to analysis of an important special case of monotone mutation operator, where the bounds obtained in the previous section become tight or asymptotically tight. In Section 5, we consider some illustrative examples of monotone mutation operators and demonstrate some applications of the general results from Section 3. In particular, in this section we obtain new lower bounds for probability to generate optimal genotypes at any given iteration t for a class of unimodal functions, for 2-SAT problem and for a family of set cover problems proposed by E. Balas (in the latter two cases we also obtain upper bounds on the runtime of the EA). Besides that, in Section 5, we give an upper bound on expected proportion of optimal genotypes for ONEMAX fitness function. Section 6 contains concluding remarks.

This work extends the conference paper (Eremeev, 2000). The extension consists in comparison of the EA behavior to that of the (1,1) EA, the (1,λ) EA, and the (1 + 1) EA in Section 3 and in the new runtime bounds and tail bounds demonstrated in Section 5. The main results from the conference paper are refined and provided with more detailed proofs.

2 Description of Algorithms and Approximating Model

2.1 Notation and Algorithms

Let the optimization problem consist in maximization of an objective function f on the set of feasible solutions Sol $\subseteq \mathcal{X} = \{0, 1\}^n$, where $\mathcal{X}$ is the search space of all binary strings of length n.

The Evolutionary Algorithm EA. The EA searches for the optimal or suboptimal solutions using a population of individuals, where each individual (genotype) g is a bitstring $(g^1, g^2, \ldots, g^n)$, and its components $g^i \in \{0, 1\}, i = 1, 2, \ldots, n$, are called genes.

In each iteration, the EA constructs a new population on the basis of the previous one. The search process is guided by the values of a fitness function

$$\phi(g) = \begin{cases} f(g) & \text{if } g \in \text{Sol}; \\ r(g) & \text{otherwise,} \end{cases}$$

where $r(\cdot)$ is a penalty function.

The individuals of the population may be ordered according to the sequence in which they are generated; thus the population may be considered as a vector of genotypes $X^t = (g_1^{(t)}, g_2^{(t)}, \ldots, g_\lambda^{(t)})$, where λ is the size of population, which is constant during the run of the EA, and t is the number of the current iteration. In this article, we consider a non-elitist algorithmic outline, where all individuals of a new population are generated independently from each other with identical probability distribution depending on the existing population only.

Each individual is generated through selection of a parent genotype by means of a selection operator, and modification of this genotype in mutation operator. During the mutation, a subset of genes in the genotype string g is randomly altered. In general, the mutation operator may be viewed as a random variable $\text{Mut}(g) \in \mathcal{X}$ with the probability distribution depending on g.

The genotypes of the initial population X^0 are generated with some a priori chosen probability distribution. The stopping criterion may be, for example, an upper bound on the number of iterations $t_{\max}$. The result is the best solution generated during the run. The EA has the following scheme:

1. Generate the initial population X^0.

2. For $t := 0$ to $t_{\max} - 1$ do

    2.1. For $k := 1$ to λ do

        Choose a parent genotype g from X^t by s-tournament selection.
        Add $g_k^{(t+1)} = \text{Mut}(g)$ to the population X^{t+1}.

In theoretical studies, the evolutionary algorithms are usually treated without a stopping criterion (see, e.g., Neumann and Witt, 2010). Unless otherwise stated, in the EA we will also assume that $t_{\max} = \infty$.

Note that in the special case of the EA with $\lambda = 1$, we can assume that $s = 1$, since the tournament selection has no effect in this case.

(1,λ) EA and (1 + 1) EA. In the following sections, we will also need a description of two simple evolutionary algorithms, known as the (1,λ) EA and the (1 + 1) EA.

The genotype of the current individual on iteration τ of the (1,λ) EA will be denoted by $b^{(\tau)}$, and in the (1 + 1) EA it will be denoted by $x^{(\tau)}$. The initial genotypes $b^{(0)}$ and $x^{(0)}$ are generated with some a priori chosen probability distribution. The only difference between the (1,λ) EA and the (1 + 1) EA consists in the method of construction of an individual for iteration $\tau + 1$ using the current individual of iteration τ as a parent. In both algorithms the new individual is built with the help of a mutation operator, which we will denote by Mut'. In the case of the (1,λ) EA, the mutation operator is independently applied λ times to the parent genotype $b^{(\tau)}$ and out of λ offspring a single genotype with the highest fitness value is chosen as $b^{(\tau+1)}$. (If there are several offspring with the highest fitness, the new individual $b^{(\tau+1)}$ is chosen arbitrarily among them.) In the (1 + 1) EA, the mutation operator is applied to $x^{(\tau)}$ once. If $x = \text{Mut}'(x^{(\tau)})$ is such that $\phi(x) > \phi(x^{(\tau)})$, then $x^{(\tau+1)} := x$; otherwise $x^{(\tau+1)} := x^{(\tau)}$.

2.2 The Proposed Model

The EA may be considered as a Markov chain in a number of ways. For example, the states of the chain may correspond to different vectors of λ genotypes that constitute the population X^t (see Rudolph, 1994). In this case, the number of states in the Markov chain is $2^{n\lambda}$. Another model representing the GA as a Markov chain is proposed in Nix
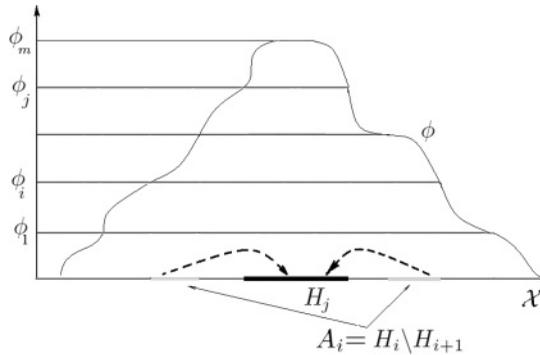
Figure 1: Transitions from A_i to H_j under mutation.

and Vose (1992), where all populations that differ only in the ordering of individuals are considered to be equivalent. Each state of this Markov chain may be represented by a vector of 2^n components, where the proportion of each genotype in the population is indicated by the corresponding coordinate and the total number of states is $\binom{2^n+\lambda-1}{\lambda}$. In the framework of this model, Vose and collaborators have obtained a number of general results concerning the emergent behavior of GAs by linking these algorithms to the infinite-population GAs (Vose, 1995).

The major difficulties in application of the above-mentioned models to the analysis of GAs for combinatorial optimization problems are connected with the necessity to use the high-grained information about fitness value of each genotype. In the present article, we consider one of the ways to avoid these difficulties by means of grouping the genotypes into larger classes on the basis of their fitness.

Assume that $\phi_0 := \min\{\phi(g) : g \in \mathcal{X}\}$ and there are m level lines of the fitness function fixed such that $\phi_0 < \phi_1 < \phi_2 \ldots < \phi_m$. The number of levels and the fitness values corresponding to them may be chosen arbitrarily, but they should be relevant to the given problem and the mutation operator to yield a meaningful model. Let us introduce the sequence of Lebesgue subsets of $\mathcal{X}$

$$H_i := \{g : \phi(g) \geq \phi_i\}, \quad i = 0, \ldots, m.$$

Obviously, $H_0 = \mathcal{X}$. For the sake of convenience, we define $H_{m+1} := \emptyset$. Also, we denote the level sets $A_i := H_i \backslash H_{i+1}$, $i = 0, \ldots, m$ which give a partition of $\mathcal{X}$. Partitioning subsets A_i are more frequently used in literature on level-based analysis, compared to the Lebesgue subsets H_i. In this article, we will frequently state that a genotype has a sufficiently high fitness; therefore, the use of subsets $H_i = \cup_{j=i}^{m} A_i$ will be more convenient in such cases. One of the partitions used in the literature, called the *canonical partition*, defines $\phi_0, \ldots, \phi_m$ as the set of all fitness values on the search space $\mathcal{X}$.

Now suppose that for all $i = 0, \ldots, m$ and $j = 1, \ldots, m$, the a priori lower bounds α_{ij} and upper bounds β_{ij} on mutation transition probabilities from subset A_i to H_j are known; that is,

$$\alpha_{ij} \leq \Pr\{\text{Mut}(g) \in H_j\} \leq \beta_{ij} \text{ for any } g \in A_i.$$

Figure 1 illustrates the transitions considered in this expression.

Let **A** denote the matrix with the elements α_{ij} where $i = 0, \ldots, m$, and $j = 1, \ldots, m$. The similar matrix of upper bounds β_{ij} is denoted by **B**. Let the population on iteration

t be represented by the *population vector*

$$\mathbf{z}^{(t)} = (z_1^{(t)}, z_2^{(t)}, \ldots, z_m^{(t)})$$

where $z_i^{(t)} \in [0, 1]$ is the proportion of genotypes from H_i in population X^t. The population vector $\mathbf{z}^{(t)}$ is a random vector, where $z_i^{(t)} \geq z_{i+1}^{(t)}$ for $i = 1, \ldots, m-1$ since $H_{i+1} \subseteq H_i$.

Let $\Pr\{g^{(t)} \in H_j\}$ be the probability that an individual, which is added after selection and mutation into X^t, has a genotype from H_j for $j = 0, \ldots, m$, and $t > 0$. According to the scheme of the EA this probability is identical for all genotypes of X^t, i.e. $\Pr\{g^{(t)} \in H_j\} = \Pr\{g_1^{(t)} \in H_j\} = \cdots = \Pr\{g_\lambda^{(t)} \in H_j\}$.

PROPOSITION 1: $\mathbf{E}[z_i^{(t)}] = \Pr\{g^{(t)} \in H_i\}$ *for all* $t > 0$, $i = 1, \ldots, m$.

PROOF: Consider the sequence of identically distributed random variables $\xi_1^i, \xi_2^i, \ldots, \xi_\lambda^i$, where $\xi_l^i = 1$ if the l-th individual in the population X^t belongs to H_i, otherwise $\xi_l^i = 0$. By the definition, $z_i^{(t)} = \sum_{l=1}^{\lambda} \xi_l^i / \lambda$, consequently $\mathbf{E}[z_i^{(t)}] = \sum_{l=1}^{\lambda} \mathbf{E}[\xi_l^i]/\lambda = \sum_{l=1}^{\lambda} \Pr\{g^{(t)} \in H_i\}/\lambda = \Pr\{g^{(t)} \in H_i\}$. □

Level-Based Mutation. If, for some mutation operator, there exist two equal matrices of lower and upper bounds $\mathbf{A}$ and $\mathbf{B}$, that is, $\alpha_{ij} = \beta_{ij}$ for all $i = 0, \ldots, m$, $j = 1, \ldots, m$ then the mutation operator will be called *level-based*. By this definition, in the case of level-based mutation, $\Pr\{\text{Mut}(g) \in H_j\}$ does not depend on a choice of genotype $g \in A_i$ and the probabilities $\gamma_{ij} = \Pr\{\text{Mut}(g) \in H_j | g \in A_i\}$ are well defined. In what follows, we call γ_{ij} a *cumulative transition probability*. The symbol $\mathbf{\Gamma}$ will denote the matrix of cumulative transition probabilities of a level-based mutation operator.

If the EA uses a level-based mutation operator, then the probability distribution of population X^{t+1} is completely determined by the vector $\mathbf{z}^{(t)}$. In this case, the EA may be viewed as a Markov chain with states corresponding to the elements of

$$Z_\lambda := \left\{ \mathbf{z} \in \{0, 1/\lambda, 2/\lambda, \ldots, 1\}^m : z_i \geq z_{i+1}, i = 1 \ldots, m-1 \right\},$$

which is the set of all possible vectors of population of size λ. Here and below, the symbol $\mathbf{z}$ is used to denote a vector from the set of all possible population vectors Z_λ.

The cardinality of set Z_λ may be evaluated analogously to the number of states in the model of Nix and Vose (1992). Now levels replace individual elements of the search space, which gives a total of $\binom{m+\lambda-1}{\lambda}$ possible population vectors.

3 Bounds on Expected Proportions of Fit Individuals

In this section, our aim is to obtain lower and upper bounds on $\mathbf{E}[\mathbf{z}^{(t)}]$ for arbitrary s and t if the distribution of the initial population is known.

Let $P_{ch}(S, \mathbf{z})$ denote the probability that the genotype, chosen by the tournament selection from a population with vector $\mathbf{z}$, belongs to a subset $S \subseteq \mathcal{X}$. Note that if the current population is represented by the vector $\mathbf{z}^{(t)} = \mathbf{z}$, then a genotype obtained by selection and mutation would belong to H_j with a conditional probability

$$\Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} = \sum_{i=0}^{m} \sum_{g \in A_i} \Pr\{\text{Mut}(g) \in H_j | g\} P_{ch}(\{g\}, \mathbf{z}). \tag{1}$$

3.1 Lower Bounds

Expression (1) and the definitions of bounds α_{ij} yield for all $j = 1, \ldots, m$:

$$\Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} \geq \sum_{i=0}^{m} \alpha_{ij} \sum_{g \in A_i} P_{ch}(\{g\}, \mathbf{z}) = \sum_{i=0}^{m} \alpha_{ij} P_{ch}(A_i, \mathbf{z}), \tag{2}$$

which turns into an equality in the case of level-based mutation and $\mathbf{A} = \mathbf{\Gamma}$.

Given a tournament size s we obtain the following selection probabilities: $P_{ch}(H_i, \mathbf{z}^{(t)}) = 1 - (1 - z_i^{(t)})^s$, $i = 1, \ldots, m$, and, consequently, $P_{ch}(A_i, \mathbf{z}) = (1 - z_{i+1})^s - (1 - z_i)^s$. This leads to the inequality:

$$\Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} \geq \sum_{i=0}^{m} \alpha_{ij}((1 - z_{i+1})^s - (1 - z_i)^s).$$

By the total probability formula,

$$\Pr\{g^{(t+1)} \in H_j\} = \sum_{\mathbf{z} \in Z_\lambda} \Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} \Pr\{\mathbf{z}^{(t)} = \mathbf{z}\} \tag{3}$$

$$\geq \sum_{\mathbf{z} \in Z_\lambda} \sum_{i=0}^{m} \alpha_{ij}((1 - z_{i+1})^s - (1 - z_i)^s) \Pr\{\mathbf{z}^{(t)} = \mathbf{z}\}$$

$$= \sum_{i=0}^{m} \alpha_{ij} \mathbf{E}[(1 - z_{i+1}^{(t)})^s - (1 - z_i^{(t)})^s]$$

$$= \alpha_{mj} \mathbf{E}[(1 - z_{m+1}^{(t)})^s] - \alpha_{0j} \mathbf{E}[(1 - z_0^{(t)})^s] - \sum_{i=1}^{m} (\alpha_{ij} - \alpha_{i-1,j}) \mathbf{E}[(1 - z_{i+1}^{(t)})^s],$$

$$\tag{4}$$

where the last expression is obtained by regrouping the summation terms. Proposition 1 implies that $\mathbf{E}[z_j^{(t+1)}] = \Pr\{g^{(t+1)} \in H_j\}$. Consequently, since $(1 - z_{m+1}^{(t)})^s = 1$ and $(1 - z_0^{(t)})^s = 0$, expression (4) gives a lower bound

$$\mathbf{E}[z_j^{(t+1)}] \geq \alpha_{mj} - \sum_{i=1}^{m} (\alpha_{ij} - \alpha_{i-1,j}) \mathbf{E}[(1 - z_i^{(t)})^s]. \tag{5}$$

Note that Eq. (5) turns into an equality in the case of level-based mutation and $\mathbf{A} = \mathbf{\Gamma}$. We would like to use Eq. (5) recursively t times in order to estimate $\mathbf{E}[\mathbf{z}^{(t)}]$ for any t, given the initial vector $\mathbf{E}[\mathbf{z}^{(0)}]$. It will be shown in the sequel that such a recursion is possible under monotonicity assumptions defined below.

Monotone Matrices and Mutation Operators. In what follows, any $((m + 1) \times m)$-matrix $\mathbf{\Delta}$ with elements δ_{ij}, $i = 0, \ldots, m$, $j = 1, \ldots, m$, will be called *monotone* iff $\delta_{i-1,j} \leq \delta_{ij}$ for all i, j from 1 to m. Monotonicity of a matrix of bounds on transition probabilities means that the greater fitness level A_i a parent solution has, the greater is its bound on transition probability to any subset H_j, $j = 1, \ldots, d$. Note that for any mutation operator, the monotone upper and lower bounds exist. Formally, for any mutation operator a valid monotone matrix of lower bounds would be $\mathbf{A} = \mathbf{0}$ where $\mathbf{0}$ is a zero matrix. A monotone matrix of upper bounds, valid for any mutation operator is $\mathbf{B} = \mathbf{U}$, where $\mathbf{U}$ is the matrix with all elements equal 1. These are extreme and impractical examples.

In reality, a problem may be connected with the absence of bounds which are sharp enough to evaluate the mutation operator properly.

If given some set of levels $\phi_1, \dots \phi_m$, there exist two matrices of lower and upper bounds $\mathbf{A}$, $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}$ and these matrices are monotone then operator Mut is called *monotone with regards to the set of levels* $\phi_1, \dots, \phi_m$. In this article, we will also call such operators *monotone* for short. Informally speaking, in the case of monotone mutation the fitter parents have a higher probability of producing fit offspring. Note that by the definition, any monotone mutation operator is level-based, since $\alpha_{ij} = \beta_{ij}$ for all i, j. The following proposition shows how the monotonicity property may be equivalently defined in terms of cumulative transition probabilities.

PROPOSITION 2: *A mutation operator* Mut *is monotone with regards to the set of levels* $\phi_1, \dots \phi_m$ *iff for any* $i, i', j \in \{0, \dots, m\}$, *such that* $i \geq i'$, *for any genotypes* $g \in A_i, g' \in A_{i'}$ *holds*

$$\Pr\{\text{Mut}(g) \in H_j\} \geq \Pr\{\text{Mut}(g') \in H_j\}.$$

PROOF: Indeed, suppose that $\mathbf{A} = \mathbf{B}$ and these matrices are monotone. Then for any genotypes $g \in A_i$ and $g' \in A_{i'}$, $i \geq i'$ holds

$$\Pr\{\text{Mut}(g) \in H_j\} \geq \alpha_{ij} \geq \alpha_{i'j} = \beta_{i'j} \geq \Pr\{\text{Mut}(g') \in H_j\}.$$

Conversely, if for any level j and any genotypes $g \in A_i$ and $g' \in A_{i'}$, $i \geq i'$ holds $\Pr\{\text{Mut}(g) \in H_j\} \geq \Pr\{\text{Mut}(g') \in H_j\}$, then taking $i = i'$ we note that $\Pr\{\text{Mut}(g) \in H_j\}$ is equal for all $g \in A_i$ and one can assign $\alpha_{ij} = \beta_{ij} = \Pr\{\text{Mut}(g) \in H_j \mid g \in A_i\}$. The resulting matrices $\mathbf{A}$ and $\mathbf{B}$ are obviously monotone. □

Proposition 2 implies that in the case of the canonical partition, that is, when $\{\phi_0, \phi_1, \dots \phi_m\}$ is the set of all values of $\phi(\cdot)$, operator Mut is monotone w.r.t. $\phi_1, \dots \phi_m$ iff for any genotypes g and g', such that $\phi(g) \geq \phi(g')$, for any $r \in \mathbb{R}$ holds

$$\Pr\{\phi(\text{Mut}(g)) \geq r\} \geq \Pr\{\phi(\text{Mut}(g')) \geq r\}.$$

The monotonicity of mutation operator with regards to a canonical partition is equivalent to the definition of monotone reproduction operator from Borisovsky and Eremeev (2001) in the case of single-parent, single-offspring reproduction. According to the terminology of Daley (1968), such random operators are also called *stochastically monotone*.

As a simple example of a monotone mutation operator we can consider a *point mutation operator*: with probability $q > 0$ keep the given genotype unchanged; otherwise (with probability $1 - q$) choose i randomly from $\{1, \dots, n\}$ and change gene i. As a fitness function we take the function OneMax $(g) \equiv \sum_{i=1}^{n} |g_i|$, where $g \in \{0, 1\}^n$. Let us assume $m = n$ and define the thresholds $\phi_0 := 0$, $\phi_1 = 1, \dots, \phi_n = n$. All genotypes with the same fitness function value have equal probability to produce an offspring with any required fitness value; therefore, this is a case of level-based mutation. In such a case, identical matrices of lower and upper bounds $\mathbf{A}$ and $\mathbf{B}$ exist and they are both equal to the matrix of cumulative transition probabilities $\mathbf{\Gamma}$. The latter consists of the following elements: $\gamma_{ij} = 1$ for all $i = 1, \dots, n$, $j = 0, \dots, i - 1$, since point mutation cannot reduce the fitness by more than one level; $\gamma_{i,i+1} = (1 - q)(n - i)/n$ for $i = 0, \dots, n - 1$ because with probability $(1 - q)(n - i)/n$ any genotype is upgraded;

$$\gamma_{ii} = \begin{cases} q + \gamma_{i,i+1} & \text{if } i = 1, \dots, n - 1; \\ q & \text{if } i = n; \end{cases}$$

because a genotype in H_i can be obtained as an offspring of a genotype from A_i in two ways: either the parent genotype has been upgraded (which happens with probability

$\gamma_{i,i+1}$) or it stays at level i, which happens with probability q; finally $\gamma_{ij} = 0$, $i = 0, \ldots, n-2$, $j = i+2, \ldots, n$ because point mutation cannot increase the level number by more than 1. The elements of matrix $\mathbf{\Gamma}$ obviously satisfy the monotonicity condition $\gamma_{ij} - \gamma_{i-1,j} \geq 0$ when $i \neq j$. For the case of $i = j$, we have $\gamma_{ii} - \gamma_{i-1,i} = q + (q-1)/n$ which is non-negative if $q \geq 1/(n+1)$. Therefore, with any $q \geq 1/(n+1)$, the matrix $\mathbf{\Gamma}$ is monotone in this example and the mutation operator is monotone as well.

PROPOSITION 3: *If $\mathbf{A}$ is monotone, then for any tournament size $s \geq 1$ and $j = 1, \ldots, m$ holds*

$$\mathbf{E}[z_j^{(t+1)}] \geq \alpha_{0j} + \sum_{i=1}^{m} (\alpha_{ij} - \alpha_{i-1,j})\mathbf{E}[z_i^{(t)}], \tag{6}$$

besides that (6) is an equality if $s = 1$, operator Mut *is monotone and $\mathbf{A}$ is its matrix of cumulative transition probabilities.*

PROOF: Monotonicity of matrix $\mathbf{A}$ implies that $\alpha_{ij} - \alpha_{i-1,j} \geq 0$ for all $i = 1, \ldots, m, j = 1, \ldots, m$, so the simple estimate $(1 - z_i^{(t)})^s \leq 1 - z_i^{(t)}$ may be applied to all terms of the sum in Eq. (5) and we get

$$\mathbf{E}[z_j^{(t+1)}] \geq \alpha_{mj} - \sum_{i=1}^{m} (\alpha_{ij} - \alpha_{i-1,j})(1 - \mathbf{E}[z_i^{(t)}]).$$

Regrouping the terms in the last bound we obtain the required Inequality (6).

Finally, note that lower bound Eq. (5) holds as an equality if the mutation operator is monotone and $\mathbf{A} = \mathbf{\Gamma}$; therefore, the last lower bound is an equality in the case of monotone $\mathbf{A} = \mathbf{\Gamma}$ and $s = 1$. □

Lower Bounds from Linear Algebra. Let $\mathbf{W}$ be a $(m \times m)$-matrix with elements $w_{ij} = \alpha_{ij} - \alpha_{i-1,j}$, let $\mathbf{I}$ be the identity matrix of the same size, and denote $\alpha = (\alpha_{01}, \ldots, \alpha_{0m})$. With these notations, Inequality (6) takes a short form $\mathbf{E}[\mathbf{z}^{(t+1)}] \geq \alpha + \mathbf{E}[\mathbf{z}^{(t)}]\mathbf{W}$. Here and below, the inequality sign "$\leq$" for some vectors $\mathbf{x} = (x_1, \ldots, x_m)$ and $\mathbf{y} = (y_1, \ldots, y_m)$ means the component-wise comparison, i.e. $\mathbf{x} \leq \mathbf{y}$ iff $x_i \leq y_i$ for all i. The following theorem gives a component-wise lower bound on vector $\mathbf{E}[\mathbf{z}^{(t+1)}]$ for any t.

THEOREM 1: *Suppose that $||\cdot||$ is some matrix norm. If matrix $\mathbf{A}$ is monotone and $\lim_{t \to \infty} ||\mathbf{W}^t|| = 0$, then for all $t \geq 1$ holds*

$$\mathbf{E}[\mathbf{z}^{(t)}] \geq \mathbf{E}[\mathbf{z}^{(0)}]\mathbf{W}^t + \alpha(\mathbf{I} - \mathbf{W})^{-1}(\mathbf{I} - \mathbf{W}^t) \tag{7}$$

and Inequality (7) turns into an equation if the tournament size $s = 1$, the mutation operator used in the EA is monotone and $\mathbf{A}$ is its matrix of cumulative transition probabilities.

The proof of this theorem is similar to the well-known inductive proof of the formula $S_t = a(1 - w)^{-1}(1 - w^t)$, $w \in \mathbb{R}$, $a \in \mathbb{R}$, for a sum of terms $a_1, \ldots, a_t$ in a geometric series $a_t = aw^{t-1}$. Note that the recursion $\mathbf{E}[\mathbf{z}^{(t+1)}] \geq \alpha + \mathbf{E}[\mathbf{z}^{(t)}]\mathbf{W}$ is similar to the recursive formula $S_{t+1} = a + S_t w$, assuming $S_0 = 0$. However, in our case, matrices and vectors replace numbers, we have to deal with inequalities rather than equalities and the initial element $\mathbf{E}[\mathbf{z}^{(0)}]$ may be nonzero unlike S_0.

PROOF OF THEOREM 1: Let us consider a sequence of m-dimensional vectors $\mathbf{u}^{(0)}, \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(t)}, \ldots$, where $\mathbf{u}^{(0)} = \mathbf{E}[\mathbf{z}^{(0)}]$, $\mathbf{u}^{(t+1)} = \alpha + \mathbf{u}^{(t)}\mathbf{W}$. We will show that $\mathbf{E}[\mathbf{z}^{(t)}] \geq \mathbf{u}^{(t)}$ for any t, using induction on t. Indeed, for $t = 0$ the inequality holds by the definition of $\mathbf{u}^{(0)}$. Now note that the right-hand side of Inequality (6) will not increase if the components of $\mathbf{E}[\mathbf{z}^{(t)}]$ are substituted with their lower bounds. Therefore, assuming we already

have $\mathbf{E}[\mathbf{z}^{(\tau)}] \geq \mathbf{u}^{(\tau)}$ for some τ and substituting $\mathbf{u}^{(\tau)}$ for $\mathbf{E}[\mathbf{z}^{(\tau)}]$ we make an inductive step $\mathbf{E}[\mathbf{z}^{(\tau+1)}] \geq \mathbf{u}^{(\tau+1)}$.

By properties of the linear operators (see, e.g., Kolmogorov and Fomin 1999, Chapter III, § 29), due to the assumption that $\lim_{t\to\infty} ||\mathbf{W}^t|| = 0$, we conclude that matrix $(\mathbf{I} - \mathbf{W})^{-1}$ exists.

Now, using the induction on t, for any $t \geq 1$ we will obtain the identity

$$\mathbf{u}^{(t)} = \mathbf{u}^{(0)}\mathbf{W}^t + \alpha(\mathbf{I} - \mathbf{W})^{-1}(\mathbf{I} - \mathbf{W}^t)$$

which leads to Inequality (7). Indeed, for the base case of $\tau = 1$, by the definition of $\mathbf{u}^{(1)}$ we have the required equality. For the inductive step, we use the following relationship

$$\mathbf{u}^{(\tau+1)} = \mathbf{u}^{(\tau)}\mathbf{W} + \alpha = \mathbf{u}^{(0)}\mathbf{W}^{\tau+1} + \alpha(\mathbf{I} - \mathbf{W})^{-1}\left(\mathbf{W} - \mathbf{W}^{\tau+1} + \mathbf{I} - \mathbf{W}\right)$$

$$= \mathbf{u}^{(0)}\mathbf{W}^{\tau+1} + \alpha(\mathbf{I} - \mathbf{W})^{-1}(\mathbf{I} - \mathbf{W}^{\tau+1}). \qquad \square$$

In conditions of Theorem 1, the right-hand side of Inequality (7) approaches $\alpha(\mathbf{I} - \mathbf{W})^{-1}$ when t tends to infinity; thus, the limit of this bound does not depend on distribution of the initial population.

In many evolutionary algorithms, an arbitrary given genotype g' may be produced with a nonzero probability as a result of mutation of any given genotype g. Suppose that the probability of such a mutation is lower bounded by some $\varepsilon > 0$ for all $g, g' \in \mathcal{X}$. Then one can obviously choose some monotone matrix $\mathbf{A}$ of lower bounds that satisfies $\alpha_{ij} \geq \varepsilon$ for all i, j. Thus, $\alpha_{mj} - \alpha_{0j} \leq 1 - \varepsilon < 1$ for all j. In this case, one can consider the matrix norm $||\mathbf{W}||_\infty = \max_j \sum_{i=1}^m |w_{ij}|$. Due to the monotonicity of $\mathbf{A}$ we have $w_{ij} = \alpha_{ij} - \alpha_{i-1,j} \geq 0$, so $||\mathbf{W}||_\infty = \max_j \sum_{i=1}^m w_{ij} = \max_j(\alpha_{mj} - \alpha_{0j}) < 1$, and the conditions of Theorem 1 are satisfied. A trivial example of a matrix that satisfies the above description would be a matrix $\mathbf{A}$ where all elements are equal to ε.

The framework of fixed budget computations, proposed in Jansen and Zarges (2012) for the RLS and the $(1+1)$ EA, naturally extends to the EAs. In this framework, one has a fixed budget of $b = (t+1)\lambda$ fitness evaluations that may be spent by the EA and it is required to estimate the expectation $\mathbf{E}[\phi_*^{(t)}]$, where $\phi_*^{(t)}$ is the greatest fitness value found during iterations $0, \ldots, t$. Theorem 1 implies m lower bounds on probabilities to generate a genotype with fitness above the specified thresholds $\phi_1, \ldots, \phi_m$ at any given iteration t: $\Pr\{g^{(t)} \in H_i\} = \mathbf{E}[z_i^{(t)}] \geq u_i^{(t)}$ where $(u_1^{(t)}, \ldots, u_m^{(t)}) = \mathbf{u}^{(t)}$ is the right-hand side of Inequality (7). These bounds may be used in the fixed budget framework: Note that $\phi_*^{(t)} \geq \phi(g_1^{(t)})$, so using the Abel transform we get a fixed budget estimate

$$\mathbf{E}[\phi_*^{(t)}] \geq \sum_{i=0}^m \Pr\{g^{(t)} \in A_i\}\phi_i = \sum_{i=0}^m (\mathbf{E}[z_i^{(t)}] - \mathbf{E}[z_{i+1}^{(t)}])\phi_i$$

$$= \phi_0\mathbf{E}[z_0^{(t)}] - \phi_m\mathbf{E}[z_{m+1}^{(t)}] + \sum_{i=1}^m (\phi_i - \phi_{i-1})\mathbf{E}[z_i^{(t)}] \geq \phi_0 + \sum_{i=1}^m (\phi_i - \phi_{i-1})u_i^{(t)}.$$

Application of Theorem 1 may be complicated due to difficulties in finding the vector $\alpha(\mathbf{I} - \mathbf{W})^{-1}$ and in estimation the effect of multiplication by matrix $\mathbf{W}^t$. Some known results from linear algebra can help to solve these tasks, as the example in Subsection 5.2 shows. However, sometimes it is possible to obtain a lower bound for $\mathbf{E}[\mathbf{z}^{(t)}]$ via analysis of the (1,1) EA algorithm, choosing an appropriate mutation operator for it. This approach is discussed below.

Lower Bounds from Associated Markov Chain. Suppose that a partition $A_0, \ldots, A_m$ defined by $\phi_0, \ldots, \phi_m$ contains no empty subsets and let $\mathbf{T}$ denote a $(m + 1) \times (m + 1)$-matrix, with components

$$t_{ij} = \alpha_{ij} - \alpha_{i,j+1}, \ i = 0, \ldots, m, \ j = 0, \ldots, m - 1,$$

$$t_{im} = \alpha_{im}, \ i = 0, \ldots, m.$$

Note that $\mathbf{T}$ is a stochastic matrix so it may be viewed as a transition matrix of a Markov chain, associated to the set of lower bounds α_{ij}. This chain is a model of the (1,1) EA, which is a special case of the (1,λ) EA with $\lambda = 1$ (see Subsection 2.1). Suppose that the (1,1) EA uses an artificial monotone mutation operator Mut′ where the cumulative transition probabilities are defined by the bounds $\alpha_{ij}, \ i = 0, \ldots, m, \ j = 1, \ldots, m$, corresponding to the EA mutation operator Mut. Namely, given a parent genotype x, for any $j = 1, \ldots, m$ we have $\Pr\{\text{Mut}'(x) \in A_j\} = \alpha_{ij} - \alpha_{i,j-1}$, where i is such that $x \in A_i$. Operator Mut′(x) may be simulated, for example, by the following two-stage procedure. At the first stage, a random index k of the offspring level is chosen with the probability distribution $\Pr\{k = j\} = \alpha_{ij} - \alpha_{i,j-1}, \ j = 1, \ldots, m$, where i is the level of parent x. At the second stage, the offspring genotype is drawn uniformly at random from A_k. (Simulation of the second stage may be computationally expensive for some fitness functions but the complexity issues are not considered now.) The initial search point $b^{(0)}$ of the (1,1) EA is generated at random with probability distribution defined by the probabilities $p_i^{(0)} := \Pr\{\xi^{(0)} \in A_i\} = \mathbf{E}[z_i^{(0)}] - \mathbf{E}[z_{i+1}^{(0)}], \ i = 0, \ldots, m$. Denoting $\mathbf{p}^{(t)} := \left(\Pr\{b^{(t)} \in A_0\}, \ldots, \Pr\{b^{(t)} \in A_m\}\right)$, by properties of Markov chains we get $\mathbf{p}^{(t)} = \mathbf{p}^{(0)} \mathbf{T}^t$. The following theorem is based on a comparison of $\mathbf{E}[z^{(t)}]$ to the distribution of the Markov chain $\mathbf{p}^{(t)}$.

THEOREM 2: *Suppose all level subsets $A_0, \ldots, A_m$ are non-empty and matrix $\mathbf{A}$ is monotone. Then for any $t = 1, 2\ldots$ holds*

$$\mathbf{E}[\mathbf{z}_i^{(t)}] \geq \mathbf{p}^{(0)} \mathbf{T}^t \mathbf{L}, \tag{8}$$

where $\mathbf{L}$ is a triangular $(m + 1) \times (m + 1)$-matrix with components $\ell_{ij} = 1$ if $i \geq j$ and $\ell_{ij} = 0$ otherwise. Besides that Inequality (8) turns into an equation if $s = 1$, the EA mutation operator is monotone and $\mathbf{A}$ is its matrix of cumulative transition probabilities.

PROOF: The (1,1) EA described above is identical to an EA′ with $\lambda = 1$, $s = 1$ and mutation operator Mut′. Let us denote the population vector of EA′ by $\hat{\mathbf{z}}^{(t)}$. Obviously,

$$\hat{z}_i^{(t)} = \sum_{k=i}^{m} \Pr\{b^{(t)} \in A_k\}, \ \ i = 1, \ldots, m. \tag{9}$$

Proposition 3 implies that in the original EA with population size λ and tournament size s, the expectation $\mathbf{E}[z^{(t)}]$ is lower bounded by the expectation $\mathbf{E}[\hat{\mathbf{z}}^{(t)}]$ since Inequality (6) holds as an equality for the whole sequence of $\mathbf{E}[\hat{\mathbf{z}}^{(t)}]$ and the right-hand side of Inequality (6) is non-decreasing on $\mathbf{E}[z_i^{(t)}]$. Equality $\mathbf{p}^{(t)} = \mathbf{p}^{(0)} \mathbf{T}^t$ together with Eq. (9) imply the required bound Inequality (8). □

Note that Inequalities (7) and (8) in Theorems 1 and 2 turn into equalities if these theorems are applied to the EA with $\lambda = 1$ and monotone mutation operator Mut′ defined above. Therefore, both theorems guarantee equal lower bounds on $\mathbf{E}[z(t)]$, given equal matrices $\mathbf{A}$.

Subsections 5.3 and 5.4 provide two examples illustrating how Theorem 2 may be used to import known results on Markov chains behavior. The example from

Subsection 5.4 employs Theorem 2 for finding a vector $\alpha(\mathbf{I} - \mathbf{W})^{-1}$, so that Theorem 1 may be applied to bound $\mathbf{E}[z_m^{(t)}]$ from below.

3.2 Upper Bounds

In this subsection, we obtain upper bounds on $\mathbf{E}[z_j^{(t+1)}]$ using a reasoning similar to the proof of Proposition 3. Expression (1) for all $j = 1, \ldots, m$ yields:

$$\Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} \le \sum_{i=0}^{m} \beta_{ij} P_{ch}(A_i, \mathbf{z}) = \sum_{i=0}^{m} \beta_{ij}((1 - z_{i+1})^s - (1 - z_i)^s), \qquad (10)$$

which turns into equality in the case of level-based mutation. By the total probability formula we have:

$$\mathbf{E}[z_j^{(t+1)}] = \sum_{\mathbf{z} \in Z_\lambda} \Pr\{g^{(t+1)} \in H_j | \mathbf{z}^{(t)} = \mathbf{z}\} \Pr\{\mathbf{z}^{(t)} = \mathbf{z}\}$$

$$\le \sum_{i=0}^{m} \beta_{ij} \mathbf{E}[(1 - z_{i+1}^{(t)})^s - (1 - z_i^{(t)})^s], \qquad (11)$$

so

$$\mathbf{E}[z_j^{(t+1)}] \le \beta_{mj} - \sum_{i=1}^{m} (\beta_{ij} - \beta_{i-1,j}) \mathbf{E}[(1 - z_i^{(t)})^s]. \qquad (12)$$

Under the expectation in the right-hand side we have a convex function on $z_i^{(t)}$. Therefore, in the case of monotone matrix $\mathbf{B}$, using Jensen's inequality (see, e.g., Rudin 1987, Chapter 3) we obtain the following proposition.

PROPOSITION 4: *If $\mathbf{B}$ is monotone then*

$$\mathbf{E}[z_j^{(t+1)}] \le \beta_{mj} - \sum_{i=1}^{m} (\beta_{ij} - \beta_{i-1,j})(1 - \mathbf{E}[z_i^{(t)}])^s. \qquad (13)$$

By means of iterative application of Inequality (13) the components of the expected population vectors $\mathbf{E}[\mathbf{z}^{(t)}]$ may be bounded up to arbitrary t, starting from the initial vector $\mathbf{E}[\mathbf{z}^{(0)}]$. The nonlinearity in the right-hand side of Inequality (13), however, creates an obstacle for obtaining an analytical result similar to the bounds of Theorems 1 and 2.

Note that all of the estimates obtained up to this point are independent of the population size and valid for arbitrary λ. In the Section 4 we will see that the right-hand side of Inequality (13) reflects the asymptotic behavior of population under monotone mutation operator as $\lambda \to \infty$.

3.3 Comparison of EA to (1,λ) EA and (1 + 1) EA

This subsection shows how the probability of generating the optimal genotypes at a given iteration of the EA relates to analogous probabilities of (1,λ) EA and (1 + 1) EA. The analysis here will be based on upper bound Inequality (13) and on some previously known results provided in the appendix.

Suppose, matrix $\mathbf{B}$ gives the upper bounds for cumulative transition probabilities of the mutation operator Mut used in the EA. Consider the (1,λ) EA and the (1 + 1) EA, based on a monotone mutation operator Mut′ for which $\mathbf{B}$ is the matrix of cumulative transition probabilities and suppose that the initial solutions $b^{(0)}$ and $x^{(0)}$ have the same distribution over the fitness levels as the best incumbent solution in the EA population X^0. Formally: $\Pr\{\text{Mut}'(x) \in H_j\} = \beta_{ij}$ for any $x \in A_i$, $i = 0, \ldots, m$, $j = 1, \ldots, m$,

and $\Pr\{b^{(0)} \in H_j\} = \Pr\{x^{(0)} \in H_j\} = \Pr\{\max_{k=1,\dots,\lambda} \phi(g_k^{(0)}) \geq \phi_j\}$, $j = 1, \dots, m$. In what follows, for any $j = 1, \dots, m$ by $P_j^{(\tau)}$ we denote the probability that current individual $b^{(\tau)}$ on iteration τ of the $(1,\lambda)$ EA belongs to H_j. Analogously $Q_j^{(\tau)}$ denotes the probability $\Pr\{x^{(\tau)} \in H_j\}$ for the $(1+1)$ EA.

The following proposition is based on upper bound Inequality (13) and the results from Borisovsky (2001) and Borisovsky and Eremeev (2001) that allow us to compare the performance of the EA, the $(1,\lambda)$ EA, and the $(1+1)$ EA.

PROPOSITION 5: *Suppose that matrix $\mathbf{B}$ is monotone. Then for any $t \geq 0$ holds*

$$\mathbf{E}[z_m^{(t+1)}] \leq \beta_{mm} - (\beta_{mm} - \beta_{m-1,m})(1 - P_m^{(t)})^s \leq \beta_{mm} - (\beta_{mm} - \beta_{m-1,m})(1 - Q_m^{(t\lambda)})^s.$$

PROOF: Let us compare the EA to the $(1,\lambda)$ EA and to the $(1+1)$ EA using the mutation and initialization procedures as described above. Theorem 6 (see the appendix) together with Proposition 1 imply that $\mathbf{E}[z_m^{(t)}] = \Pr\{g^{(t)} \in H_m\} \leq P_m^{(t)}$ for all $t \geq 0$. Furthermore, Theorem 5 from Borisovsky and Eremeev (2001) (see the appendix) implies that $P_m^{(t)} \leq Q_m^{(t\lambda)}$ for all $t \geq 0$. Using Proposition 4 and monotonicity of $\mathbf{B}$, we conclude that both claimed inequalities hold. $\qquad\square$

4 EA with Monotone Mutation Operator

First of all, note that in the case of the monotone mutation operator, two equal monotone matrices of lower and upper bounds $\mathbf{A} = \mathbf{B}$ exist, so the bounds Eq. (5) and Eq. (12) give equal results, and assuming $\mathbf{\Gamma} = \mathbf{A} = \mathbf{B}$ we get

$$\mathbf{E}[z_j^{(t+1)}] = \gamma_{mj} - \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j})\mathbf{E}[(1 - z_i^{(t)})^s], \quad j = 1, \dots, m, \ \ t = 0, 1, \dots. \tag{14}$$

This equality will be used several times in what follows.

In general, the population vectors are random values whose distributions depend on λ. To express this in the notation, let us denote the proportion of genotypes from H_i in population X^t by $z_i^{(t)}(\lambda)$, $i = 1, \dots, m$.

The following Lemma 1 and Theorem 3 based on this lemma indicate that in the case of monotone mutation, recursive application of the formula from right-hand side of upper bound Eq. (13) allows to compute the expected population vector of the infinite-population EA at any iteration t.

LEMMA 1: *Let the EA use a monotone mutation operator with cumulative transition probabilities matrix $\mathbf{\Gamma}$, and let the genotypes of the initial population be identically distributed. Then*

(i) *for all $t = 0, 1, \dots$ and $i = 1, \dots, m$ holds*

$$\lim_{\lambda \to \infty} \left(\mathbf{E}\left[(1 - z_i^{(t)}(\lambda))^s\right] - \left(1 - \mathbf{E}[z_i^{(t)}(\lambda)]\right)^s\right) = 0; \tag{15}$$

(ii) *if the sequence of m-dimensional vectors $\mathbf{u}^{(0)}, \mathbf{u}^{(1)}, \dots, \mathbf{u}^{(t)}, \dots$ is defined as*

$$\mathbf{u}^{(0)} = \mathbf{E}[\mathbf{z}^{(0)}(\lambda)], \tag{16}$$

$$u_j^{(t+1)} = \gamma_{mj} - \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j})(1 - u_i^{(t)})^s \tag{17}$$

for $j = 1, \ldots, m$ and $t \geq 0$. Then $\lim_{\lambda \to \infty} \mathbf{E}[\mathbf{z}^{(t)}(\lambda)] = \mathbf{u}^{(t)}$ for all $j = 1, \ldots, m$ at any iteration t.

The main step in the proof of Lemma 1 (i) will consist in showing that for a supplementary random variable $X = (1 - z_i^{(t)}(\lambda))^s - (1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s$, the value of $|\mathbf{E}[X]|$ is upper-bounded by an arbitrary small $\varepsilon > 0$. This step is made by splitting the range $[-1, 1]$ of X into a "high-probability" area and a "low-probability" area in such a way that $|X|$ is at most ε in the "high-probability" area. An analogous technique is used, for example, in the proof of Lebesgue Theorem (see, e.g., Kolmogorov and Fomin, 1999, Chapter VII, Section 44).

PROOF OF LEMMA 1: From Eq. (14), we conclude that if statement (i) holds, then with $\lambda \to \infty$, the convergence of $\mathbf{E}[\mathbf{z}^{(t)}(\lambda)]$ to $\mathbf{u}^{(t)}$ will imply that $\mathbf{E}[\mathbf{z}^{(t+1)}(\lambda)] \to \mathbf{u}^{(t+1)}$. Thus, statement (ii) follows by induction on t.

Let us now prove statement (i). Given some t, to prove Eq. (15) we recall the sequence of i.i.d. random variables $\mathcal{I}_1^i, \mathcal{I}_2^i, \ldots, \mathcal{I}_\lambda^i$, where $\mathcal{I}_k^i = 1$, if the k-th individual of population X^t belongs to H_i, otherwise $\mathcal{I}_k^i = 0$. By the law of large numbers, for any $i = 1, \ldots, m$ and $\varepsilon > 0$, we have

$$\lim_{\lambda \to \infty} \Pr \left\{ \left| \frac{\sum_{k=1}^\lambda \mathcal{I}_k^i}{\lambda} - \mathbf{E}[\mathcal{I}_1^i] \right| < \varepsilon \right\} = 1.$$

Note that $\sum_{k=1}^\lambda \mathcal{I}_k^i / \lambda = z_i^{(t)}(\lambda)$. Besides that, due to Proposition 1, $\mathbf{E}[\mathcal{I}_1^i] = \Pr\{\mathcal{I}_1^i = 1\} = \mathbf{E}[z_i^{(t)}(\lambda)]$. (In the case of $t = 0$ this equality holds as well, since all individuals of the initial population are distributed identically.) Therefore, for any $\varepsilon > 0$ the convergence $\Pr\left\{\left|z_i^{(t)}(\lambda) - \mathbf{E}[z_i^{(t)}(\lambda)]\right| < \varepsilon\right\} \longrightarrow 1$ holds. Now by continuity of the function $(1 - x)^s$, it follows that

$$\lim_{\lambda \to \infty} \Pr \left\{ \left| (1 - z_i^{(t)}(\lambda))^s - (1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s \right| \geq \varepsilon \right\} = 0.$$

Let us denote $F_\lambda(x) := \Pr\left\{(1 - z_i^{(t)}(\lambda))^s - (1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s < x\right\}$. Then

$$\lim_{\lambda \to \infty} \left(\mathbf{E}\left[(1 - z_i^{(t)}(\lambda))^s\right] - (1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s \right) = \lim_{\lambda \to \infty} \int_{-1}^1 x \, dF_\lambda(x).$$

For arbitrary $\varepsilon > 0$ we can split the integration domain into two subsets $[-1, +1]\setminus[-\varepsilon, \varepsilon]$ and $[-\varepsilon, \varepsilon]$ and bound x by 1 in the first case and by ε in the second case. Then in view of the definition of $dF_\lambda(x)$ we get

$$\int_{-1}^1 x \, dF_\lambda(x) \leq \Pr \left\{ \left| (1 - z_i^{(t)}(\lambda))^s - (1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s \right| \geq \varepsilon \right\} + \int_{|x|<\varepsilon} \varepsilon \, dF_\lambda(x).$$

The first term tends to 0 as $\lambda \to \infty$. The second term is at most ε for any $\varepsilon > 0$. Hence Eq. (15) holds. □

Combining Equality (14) with claim (i) of Lemma 1 we obtain a recursive expression for $\mathbf{E}[\mathbf{z}^{(t)}]$ in the infinite-population EA, which is formulated as

THEOREM 3: *If the mutation operator is monotone and individuals of the initial population are distributed identically, then*

$$\lim_{\lambda \to \infty} \mathbf{E}[z_j^{(t+1)}(\lambda)] = \gamma_{mj} - \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j})(1 - \mathbf{E}[z_i^{(t)}(\lambda)])^s \qquad (18)$$

for all $j = 1, \ldots, m, \ t \geq 0$.

For any i, j and $t > 0$, the term $u_j^{(t)}$ of the sequence defined by Eq. (17) is nondecreasing in $u_i^{(t-1)}$ and in s as well. With this in mind, we can expect that the components of population vector of the infinite-population EA will typically increase with the tournament size. Theorem 4 below gives a rigorous proof of this fact under some technical conditions on distributions of Mut and X^0.

THEOREM 4: *Let* $\mathbf{z}^{(t)}$ *and* $\hat{\mathbf{z}}^{(t)}$ *correspond to EAs with tournament sizes* s *and* $\hat{s}$, *where* $s < \hat{s}$. *Besides that, suppose that* Mut *is monotone with* $\gamma_{mj} > \gamma_{0j}$ *for all* $j = 1, \ldots, m$ *and the individuals of initial populations are identically distributed so that* $\Pr\{g^{(0)} \in H_i\} \in (0, 1)$ *for all* $i = 1, \ldots, m$. *Then for any* $t > 0$, *given a sufficiently large* λ, *holds*

$$\mathbf{E}[\hat{z}_i^{(t)}(\lambda)] > \mathbf{E}[z_i^{(t)}(\lambda)], \quad i = 1, \ldots, m.$$

PROOF: Let the sequences $\{\mathbf{u}^{(t)}\}$ and $\{\hat{\mathbf{u}}^{(t)}\}$ be defined as in Lemma 1, corresponding to tournament sizes s and $\hat{s}$. By the above assumptions, $\mathbf{u}^{(0)} = \hat{\mathbf{u}}^{(0)}$.

Now since $\Pr\{g^{(0)} \in H_i\} \in (0, 1)$ for all $i = 1, \ldots, m$, we have $u_i^{(0)} = \hat{u}_i^{(0)} \in (0, 1)$ for any $i = 1, \ldots, m$. Thus, for all $j = 1, \ldots, m$ holds

$$u_j^{(1)} = \gamma_{mj} - \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j})(1 - u_i^{(0)})^s < \gamma_{mj} - \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j})(1 - \hat{u}_i^{(0)})^{\hat{s}} = \hat{u}_j^{(1)}, \quad (19)$$

since $s < \hat{s}$ and $\gamma_{ij} - \gamma_{i-1,j} > 0$ at least for one of the levels i according to the assumption that $\gamma_{mj} > \gamma_{0j}$. Due to the same reason, for all $j = 1, \ldots, m$ from the last equality in Eq. (19) we get $\hat{u}_j^{(1)} < \gamma_{mj} \leq 1$. Using the fact that $(1 - u_i^{(0)})^s \leq 1 - u_i^{(0)}$ and rearranging the terms as in the proof of Proposition 3 we get

$$u_j^{(1)} \geq \gamma_{0j} + \sum_{i=1}^{m}(\gamma_{ij} - \gamma_{i-1,j}) \, u_i^{(0)} > 0.$$

To sum up, for $t = 1$ we have $u_i^{(1)} < \hat{u}_i^{(1)}, u_i^{(1)} \in (0, 1)$ and $\hat{u}_i^{(1)} \in (0, 1)$.

Furthermore, if we assume that for all $i = 1, \ldots, m$ holds $u_i^{(t-1)} < \hat{u}_i^{(t-1)}$, $u_i^{(t-1)} \in (0, 1)$ and $\hat{u}_i^{(t-1)} \in (0, 1)$ then analogously to Eq. (19) we get $u_j^{(t)} < \hat{u}_j^{(t)}$ for all $j = 1, \ldots, m$. Besides that, just as in the case of $t = 1$ we get $\hat{u}_j^{(t)} \in (0, 1)$ and $u_j^{(t)} \in (0, 1)$. So by induction we conclude that $u_j^{(t)} < \hat{u}_j^{(t)}$ for all $j = 1, \ldots, m$ and all $t > 0$.

Finally, by claim (ii) of Lemma 1, for any i and t, given a sufficiently large λ, holds $\mathbf{E}[\hat{z}_i^{(t)}(\lambda)] > \mathbf{E}[z_i^{(t)}(\lambda)]$. □

Informally speaking, Theorem 4 implies that in the case of the monotone mutation operator an optimal selection mechanism consists in setting $s \to \infty$, which actually converts the EA into the $(1, \lambda)$ EA.

5   Applications and Illustrative Examples

5.1   Examples of Monotone Mutation Operators

Let us consider two cases where the mutation is monotone and the matrices $\boldsymbol{\Gamma}$ have a similar form.

First, we consider the simple fitness function OneMax (g). Suppose that the EA uses the bitwise mutation operator, changing every gene with a given probability p_m, independently of the other genes. Let the subsets $H_0, \ldots, H_m$ be defined by the level lines $\phi_0 = 0, \phi_1 = 1, \ldots, \phi_m = m$ and $m = n$. The matrix $\boldsymbol{\Gamma}$ for this operator could be obtained using the result from Bäck (1992), but here we shall consider this example as a special case of a more general setting.

Let the representation of the problem admit a decomposition of the genotype string into d nonoverlapping substrings (called *blocks* here) in such a way that the fitness function equals the number of blocks for which a certain property $\mathcal{K}$ holds. The functions of this type belong to the class of additively decomposed functions, where the elementary functions are Boolean and substrings are non-overlapping (see, e.g., Mühlenbein et al., 1999). Let $K(g, \ell) = 1$ if $\mathcal{K}$ holds for the block ℓ of genotype g, and $K(g, \ell) = 0$ otherwise (here $\ell = 1, \ldots, d$).

Suppose that during mutation, any block for which $\mathcal{K}$ did not hold, gets the property $\mathcal{K}$ with probability $\tilde{r}$, that is,

$$\Pr\{K(\mathrm{Mut}(g), \ell) = 1 | K(g, \ell) = 0\} = \tilde{r}, \quad \ell = 1, \ldots, d.$$

On the other hand, assume that a block with the property $\mathcal{K}$ keeps this property during mutation with probability r, that is,

$$\Pr\{K(\mathrm{Mut}(g), \ell) = 1 | K(g, \ell) = 1\} = r, \quad \ell = 1, \ldots, m.$$

Let $m = d$ and the subsets $H_0, \ldots, H_m$ correspond to the level lines $\phi_0 = 0, \phi_1 = 1, \ldots, \phi_m = m$ again. In this case, the element γ_{ij} of cumulative transition probabilities matrix $\boldsymbol{\Gamma}$ equals the probability to obtain a genotype containing j or more blocks with property $\mathcal{K}$ after mutation of a genotype which contained i blocks with this property. Let $P(k', k)$ denote the probability that during mutation k' blocks without property $\mathcal{K}$ would produce k blocks with this property and let $Q(i, l)$ denote the probability that after mutation of a set of i blocks with property $\mathcal{K}$, there will be at least l blocks with property $\mathcal{K}$ among them. (If $l > i$ then $Q(i, l) := 0$.) With these notations,

$$\gamma_{ij} = \sum_{k=0}^{m-i} P(m - i, k) Q(i, j - k).$$

Clearly, $P(k', k) = \binom{k'}{k} \tilde{r}^k (1 - \tilde{r})^{k'-k}$ and $Q(i, l) = \sum_{v=0}^{\min\{i, i-l\}} \binom{i}{v} (1 - r)^v r^{i-v}$. Thus,

$$\gamma_{ij} = \sum_{k=0}^{m-i} \binom{m-i}{k} \tilde{r}^k (1 - \tilde{r})^{m-i-k} \sum_{v=0}^{\min\{i, i-(j-k)\}} \binom{i}{v} (1 - r)^v r^{i-v}. \tag{20}$$

It is shown in Eremeev (2000) and Borisovsky and Eremeev (2008) that if $r \geq \tilde{r}$, then matrix $\boldsymbol{\Gamma}$ defined by Expression (20) is monotone.

Now matrix $\boldsymbol{\Gamma}$ for the bitwise mutation on OneMax function is obtained assuming that $\tilde{r} = (1 - r) = p_\mathrm{m}$ and $m = d = n$. This operator is monotone in view of the above mentioned result, if $p_\mathrm{m} \leq 0.5$, since in this case $r \geq \tilde{r}$. The monotonicity of bitwise mutation on OneMax is used in works of Doerr et al. (2010) and Witt (2013).

Expression (20) may be also used for finding the cumulative transition matrices of some other optimization problems with a regular structure. As an example, below we consider the vertex cover problem (VCP) on graphs of a special structure.

In general, the vertex cover problem is formulated as follows. Let $G = (V, E)$ be a graph with a set of vertices $V = \{v_1, \dots, v_{|V|}\}$ and the edge set $E = \{e_1, \dots, e_{|E|}\}$ where $e_i = \{u(i), v(i)\} \subseteq V, \ i = 1, \dots, |E|$. A subset $C \subseteq V$ is called a vertex cover of G if every edge has at least one endpoint in C. The vertex cover problem is to find a vertex cover C^* of minimal cardinality.

Suppose that the VCP is handled by the EA with the following representation: each gene $g^i \in \{0, 1\}, i = 1, \dots, |E|$ corresponds to an edge e_i of G, assigning one of its endpoints which has to be included in the cover $C(g)$. To be specific, we can assume that $g^i = 1$ means that $u(i) \in C(g)$ and $g^i = 0$ means that $v(i) \in C(g)$. The vertices, not assigned by one of the chosen endpoints, do not belong to $C(g)$. On one hand, this edge-based representation is degenerate in the sense that one vertex cover C may be encoded by different genotypes g. On the other hand, any genotype g defines a feasible cover $C(g)$. A natural way to choose the fitness function in the case of this representation is to assume $\phi(g) = |V| - |C(g)|$.

Note that most publications on evolutionary algorithms for VCP use the vertex-based representation with $|V|$ genes, where $g_j = 1, \ j = 1, \dots, |V|$ implies inclusion of vertex v_j into C (see, e.g., Neumann and Witt, 2010, § 12.1). In contrast to the edge-based representation, the vertex-based representation is not degenerate but some genotypes in this representation may define infeasible solutions.

Following Saiko (1989) we denote by $G(m)$ the graph consisting of m disconnected triangle subgraphs. Each triangle is covered optimally by two vertices and the redundant cover consists of three vertices. In spite of simplicity of this problem, it is proven in Saiko (1989) that some well-known algorithms of branch and bound type require exponential in m number of iterations if applied to the VCP on graph $G(m)$.

In the case of $G(m)$, the fitness $\phi(g)$ coincides with the number of optimally covered triangles in $C(g)$ (i.e., triangles where only two different vertices are chosen), since covering nonoptimally all triangles gives $C(g) = V$ and each optimally covered triangle decreases the size of the cover by one. Let the genes representing the same triangle constitute a single block, and let the property $\mathcal{K}$ imply that a triangle is optimally covered. Then by looking at the two possible ways to produce a gene triplet that redundantly covers a triangle, (i) given a redundant triangle and (ii) given an optimally covered triangle, we conclude that (i) $\tilde{r} = 1 - p_m^3 - (1 - p_m)^3$ and (ii) $r = 1 - p_m(1 - p_m)^2 - p_m^2(1 - p_m)$. Using Expression (20) we obtain the cumulative transition matrix for this mutation operator. It is easy to verify that in this case the inequality $r \geq \tilde{r}$ holds for any mutation probability p_m, and therefore the operator is always monotone.

Computational Experiments. Below we present some experimental results in comparison with the theoretical estimates obtained in Section 3. To this end we consider an application of the EA to the VCP on graphs $G(m)$. The average proportion of optimal genotypes in the population for different population sizes is presented in Figure 2. Here $m = 8, p_m = 0.1, s = 2$, and $\mathbf{z}^{(0)} = \mathbf{0}$ (these parameters are chosen to ensure clear visibility on plots). The statistics is accumulated in 1000 independent runs of the algorithm where for each t only one individual $g_1^{(t)}$ was checked for optimality. Thus for each t we have a series of 1000 Bernoulli trials with a success probability $\Pr\{g_1^{(t)} \in H_m\} = \mathbf{E}[z_m^{(t)}]$ which is estimated from the experimental data. The 95%-confidence intervals for success probability in Bernoulli trials are computed using the Normal approximation as described in Cramer (1946, Chapter 34).
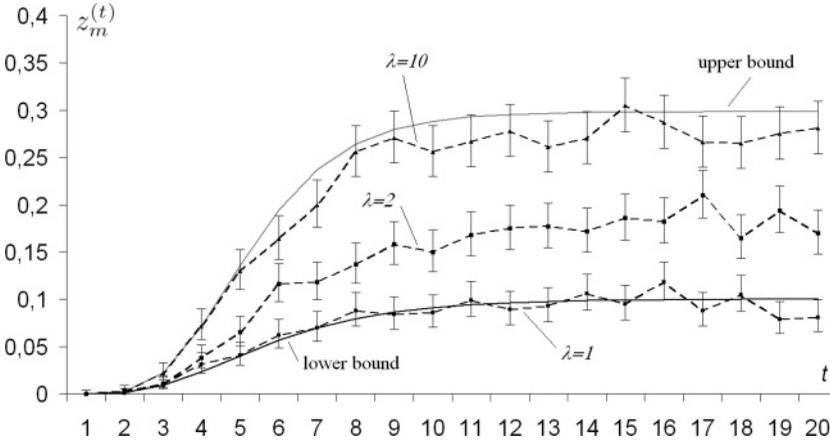
Figure 2: Average proportion of optimal VCP solutions and the theoretical lower and upper bounds as functions of the iteration number. Here $s = 2$, $\lambda = 1, 2$, and 10.
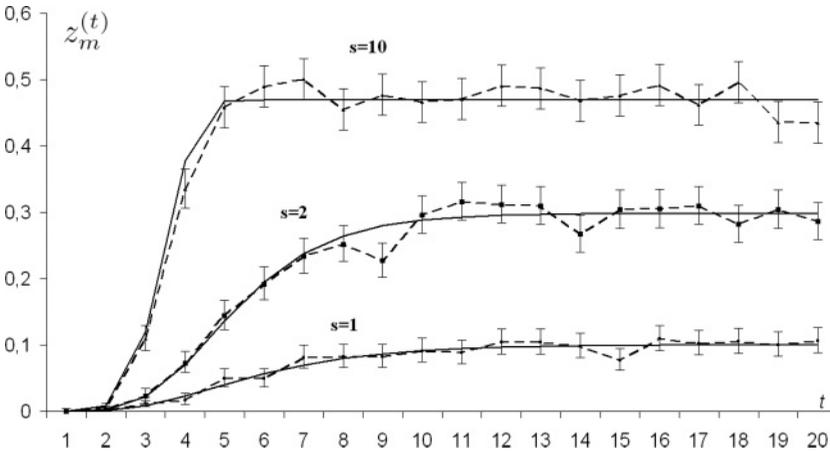

Figure 3: Average proportion of optimal solutions to VCP and the theoretical upper bound, as functions of the iteration number. Here $\lambda = 100$, $s = 1, 2$, and 10.

The experimental results are shown in dashed lines. The solid lines correspond to the lower and upper bounds given by Expressions (7) and (13). The plot shows that upper bound Expression (13) gives a good approximation to the value of $z_m^{(t)}$ even if the population size is not large. The lower bound Expression (7) coincides with the experimental results when $\lambda = 1$, up to a minor sampling error.

Another series of experiments was carried out to compare the behavior of EAs with different tournament sizes. Figure 3 presents the experimental results for 1000 runs of the EA with $p_m = 0.1$, $\lambda = 100$ and $\mathbf{z}^{(0)} = \mathbf{0}$ solving the VCP on $G(8)$. This plot demonstrates the increase in the average proportion of the optimal genotypes as a function of the tournament size, which is consistent with Theorem 4. The 95%-confidence intervals are found as described above.

5.2   Lower Bound for Randomized Local Search on Unimodal Functions.

First of all, let us describe the RLS algorithm which will be implicitly studied in this subsection. At each iteration of RLS the current genotype x is stored. In the beginning of RLS execution, x is initialized with some probability distribution (e.g., uniformly over $\mathcal{X}$). An iteration of RLS consists in building an offspring y of x by flipping exactly one randomly chosen bit in x. If $\phi(y) \geq \phi(x)$, then x is replaced by the new genotype y. The process continues until some termination condition is met.

Below we will illustrate the usage of Theorem 1 on the class of ℓ-Unimodal functions. In this class, each function has exactly ℓ distinctive fitness values $\phi_0 < \phi_1 < \cdots < \phi_{\ell-1}$, and each solution in the search space is either optimal or its fitness may be improved by flipping a single bit. Naturally we assume that $m = \ell - 1$ and that level A_m consists of optimal solutions.

As a mutation operator in the EA we will use a routine denoted by $\mathrm{Mut_{RLS}}$: given a genotype g, this routine first changes one randomly chosen gene and if this modification improves the genotype fitness, then $\mathrm{Mut_{RLS}}$ outputs the modified genotype, otherwise $\mathrm{Mut_{RLS}}(g)$ outputs the genotype g unchanged. Note that in the case of $\lambda = 1$, the EA with $\mathrm{Mut_{RLS}}$ mutation becomes a version of RLS. The lower bounds from Section 3 are tight for $\lambda = 1$ (which implies $s = 1$), therefore the following analysis in this subsection may be viewed primarily as a study of the RLS.

Mutation operator $\mathrm{Mut_{RLS}}$ never decreases the genotype fitness and improves any non-optimal genotype with probability at least $1/n$, so we have $\alpha_{ij} = 1$ for all $i = 1, \ldots, m, \ j = 0, \ldots, i$ and $\alpha_{i,i+1} = 1/n$ for $i = 0, \ldots, m-1$. The chances for improvements by more that one fitness level are not foreseeable, so we put $\alpha_{ij} = 0$ for all $i = 0, \ldots, m-2, \ j = i + 2, \ldots, m$. Note that this matrix $\mathbf{A}$ is monotone.

Now $\alpha = (1/n, 0, \ldots, 0)$ and the matrix $\mathbf{W}$ consists of the following elements:

$$w_{ij} = \alpha_{ij} - \alpha_{i-1,j} = \begin{cases} 1/n & \text{if } i = j + 1; \\ 1 - 1/n & \text{if } i = j. \\ 0 & \text{otherwise} \end{cases}.$$

In order to apply Theorem 1 we also need to choose an appropriate matrix norm and evaluate this norm for matrix $\mathbf{W}$. In this particular application we will use $|| \cdot ||_2$, which is the matrix norm induced by the Euclidean vector norm in $\mathbb{R}^m$. It is well-known that for any matrix $\mathbf{W}$ holds $||\mathbf{W}||_2 = \sqrt{\lambda_{\max}}$, where $\lambda_{\max}$ is the maximal eigenvalue of matrix $\mathbf{W}\mathbf{W}^T$ (see, e.g., Section 9.5 in Collatz, 1966). Here and below $\mathbf{W}^T$ denotes the transpose of matrix $\mathbf{W}$.

It is easy to check that matrix $\mathbf{W}\mathbf{W}^T$ is composed of zero elements everywhere except for m diagonal elements, $m - 1$ superdiagonal and $m - 1$ subdiagonal elements. In particular, it has identical elements $(1 + (n - 1)^2)/n^2$ on the diagonal and all superdiagonal and subdiagonal elements are equal to $(n - 1)/n^2$. This matrix $\mathbf{W}\mathbf{W}^T$ belongs to the class of tridiagonal Toeplitz matrices and its maximal eigenvalue is

$$\lambda_{\max} = \frac{1 + (n - 1)^2}{n^2} + \frac{2(n - 1)}{n^2} \cos \frac{\pi}{\ell}.$$

(see Theorem 7 in the appendix). Therefore

$$||\mathbf{W}||_2 = \sqrt{1 - \frac{2(n - 1)}{n^2} \left(1 - \cos \frac{\pi}{\ell} \right)}.$$

So $||\mathbf{W}||_2 < 1$ and since matrix $\mathbf{A}$ is monotone, we can apply Theorem 1.

Let us denote $\mathbf{e} := (1, 1, \ldots, 1) \in \mathbb{R}^m$. The vector $\mathbf{v} = \mathbf{e}$ satisfies the equation $\mathbf{v} = \alpha(\mathbf{I} - \mathbf{W})^{-1}$ and since $||\mathbf{W}||_2 < 1$, the right-hand side in Inequality (7) of Theorem 1 tends to $\mathbf{e}$ as $t \to \infty$.

In order to obtain an explicit lower bound on $\mathbf{E}[z_m^{(t)}]$ for any given t, we will evaluate the speed of convergence of the right-hand side in Inequality (7) to $\mathbf{e}$. Note that by properties of matrix norms we have

$$||\mathbf{e}\mathbf{W}^t||_2 \leq ||\mathbf{e}||_2 \cdot ||\mathbf{W}||_2^t = \sqrt{m}\,||\mathbf{W}||_2^t. \tag{21}$$

Thus for any distribution of initial population Theorem 1 gives a lower bound

$$\mathbf{E}[\mathbf{z}^{(t)}] \geq \mathbf{e}(\mathbf{I} - \mathbf{W}^t) \geq \mathbf{e} - \sqrt{m}\,||\mathbf{W}||_2^t \cdot \mathbf{e},$$

where the last inequality holds because each component of vector $\mathbf{e}\mathbf{W}^t$ is upper-bounded by $||\mathbf{e}\mathbf{W}^t||_2$ which is at most $\sqrt{m}\,||\mathbf{W}||_2^t$ by Inequality (21).

Finally, independently of population size λ and tournament size s we get a lower bound for the proportion of optimal genotypes in the EA population:

$$\mathbf{E}[z_m^{(t)}] \geq 1 - \sqrt{\ell - 1}\left(1 - \frac{2(n-1)}{n^2}\left(1 - \cos\frac{\pi}{\ell}\right)\right)^{t/2}. \tag{22}$$

The Taylor expansion for $\cos(x)$ gives

$$\cos\frac{\pi}{\ell} \leq 1 - \frac{\pi^2}{2\ell^2} + \frac{\pi^4}{24\ell^4} \leq 1 - \frac{\pi^2(\ell-1)^2}{2\ell^4}.$$

Now since $\sqrt{1-x} \leq 1 - x/2$ and $\ln(1-x) \leq -x$, we obtain

$$\mathbf{E}[z_m^{(t)}] \geq 1 - \sqrt{\ell-1}\left(1 - \frac{\pi^2(n-1)(\ell-1)^2}{2\ell^4 n^2}\right)^t \geq 1 - \exp\left\{\frac{\ln(\ell-1)}{2} - \frac{t\pi^2}{\ell^2 n}\left(1 - \frac{2}{\ell}\right)\right\}.$$

In the case of RLS, that is, when $\lambda = 1$, this gives the following tail bound

COROLLARY 1: *The probability that the maximum of a fitness function from ℓ-UNIMODAL is first reached after more than t iterations of RLS is at most* $\sqrt{e}^{\ln(\ell-1) - t\ell^{-2}n^{-1}(\pi^2 - 20\ell^{-1})}$.

A positive feature of this tail bound is that it approaches to 0 exponentially fast in t. A weakness of Corollary 1 is that its bound is greater than 1 (and therefore useless) when $t < \ln(\ell-1)\ell^2 n/(\pi^2 - 20\ell^{-1})$. The obtained tail bound may be improved for some relatively small t using the expected RLS runtime bound and Markov inequality. Let T denote the number of fitness evaluations made in RLS until the optimum is achieved. Then the RLS runtime $\mathbf{E}[T] \leq n(\ell-1)$ since each fitness level requires on average at most n iterations of RLS. By Markov inequality we have $\Pr\{T \geq t\} \leq n(\ell-1)/t$. This tail bound becomes meaningful as soon as t reaches $n(\ell-1)$ but it does not give an exponential convergence and therefore yields to Corollary 1 for large t. It would be interesting to compare our tail bounds to those obtainable by the approach from Lehre and Witt (2014) but tight analysis of RLS is beyond the scope of this article.

5.3 Lower Bounds and Runtime Analysis for the 2-SAT Problem

The Satisfiability problem (SAT) in general is known to be NP-complete (Garey and Johnson, 1979), but it is polynomially solvable in the special case denoted by the 2-SAT: given a Boolean formula with CNF where each clause contains at most two literals, find out whether a satisfying assignment of variables exists.

Let n be the number of logical variables and let m be the number of clauses in the CNF. A natural encoding of solutions is a binary string g where $g_i = 1$ if the i-th logical variable has the value "true" and otherwise $g_i = 0$.

We consider an EA with the tournament size $s = 1$ and the following mutation operator Mut$_{SAT}$: Draw randomly a clause which is not satisfied, choose one variable among the variables of the clause at random, and modify this variable. Otherwise, keep the solution unchanged. This method of random perturbation was proposed in the randomized algorithm of Papadimitriou (1991) for 2-SAT, which has the runtime $O(n^2)$, if the CNF is satisfiable. A generalization of the algorithm from Papadimitriou (1991) to the general case of SAT, known as WalkSat algorithm, shows competitive experimental results (Selman et al., 1996). In the special case of SAT, where each clause contains at most k literals, which is denoted by k-SAT, algorithm WalkSat has a runtime bound $O((2 - 2/k)^k)$ (Schöning, 1999).

A fitness function does not influence the EA execution when $s = 1$ but it will be useful for our theoretical analysis. Let us assume that $\phi(g)$ equals the Hamming distance to a satisfying assignment g^*. Here and below, we assume that at least one satisfying assignment g^* exists.

For any nonsatisfying truth assignment, the fitness improvement probability is $1/2$. So we can apply the following monotone bounds: $\alpha_{ij} = 1$ for all $i = 1, \dots, m$, $j = 0, \dots, i - 1$; $\alpha_{i,i+1} = 1/2$ for $i = 0, \dots, m - 1$;

$$\alpha_{ii} = \begin{cases} 1/2 & \text{if } i = 1, \dots, m - 1; \\ 1 & \text{if } i = m; \end{cases}$$

$\alpha_{ij} = 0$, $i = 0, \dots, m - 2$, $j = i + 2, \dots, m$. These lower bounds define the Markov chain transition probabilities $\mathbf{T}$ with $t_{ij} = \alpha_{ij} - \alpha_{i,j+1}$, $i = 0, \dots, m$, $j = 0, \dots, m - 1$, and $t_{im} = \alpha_{im}$, $i = 0, \dots, m$ according to Subsection 3.1. It turns out that this matrix $\mathbf{T}$ is the same as the transition matrix of the symmetric Gambler's Ruin random walk with one reflecting barrier (state 0) and one absorbing barrier (state m): $t_{0,1} = 1$, $t_{i,i+1} = t_{i,i-1} = 1/2$ for $i = 1, \dots, m - 1$, $t_{mm} = 1$, all other elements t_{ij} are equal to zero. The result from Papadimitriou (1991) implies that, regardless of the initial state, there exists a constant $c > 0$, such that after cn^2 transitions the absorbing probability of this random walk is at least $1/2$. This means that $p_m^{(cn^2)} \geq 1/2$ and the m-th component of the vector $\mathbf{p}^{(0)} \mathbf{T}^t \mathbf{L}$ is at least $1/2$ as well. Therefore, Theorem 2 yields

COROLLARY 2: *If the EA for 2-SAT has the tournament size $s = 1$ and the mutation operator* Mut$_{SAT}$, *then the probability to generate a satisfying assignment in population X^{cn^2} is at least $1/2$ for some constant $c > 0$.*

It makes sense to apply Theorem 2 only in the case of $s = 1$ in this example, since for $s > 1$ the tournament selection is impossible without computing the Hamming distance to a satisfying assignment which is unknown.

If the EA with $s = 1$ and mutation Mut$_{SAT}$ is restarted every t_{max} iterations and $t_{max} = cn^2$, then the overall runtime of this iterated EA is $O(\lambda n^2)$ by Corollary 2 and Markov inequality. Note that Corollary 2 holds for any distribution of the initial population, so the runtime bound $O(\lambda n^2)$ applies to the EA without restarts as well. In a similar way the EA with Mut$_{SAT}$ can simulate the randomized algorithm of Schöning (1999) for k-SAT with runtime $O((2 - 2/k)^k)$.

5.4 Lower Bounds and Runtime Analysis for Balas Set Cover Problems

In general, the set cover problem (SCP) is formulated as follows. Given: a ground set M and a set of covering subsets $M_j \subseteq M$, with indices $j \in U := \{1, \dots, n\}$. A subset of indices $J \subseteq U$ is called a *cover* if $\cup_{j \in J} M_j = M$. The goal is to find a cover of minimum cardinality. In what follows, for any $i \in M$ we denote by N_i the set of numbers of the

subsets that cover an element i, i.e. $N_i = \{j : i \in M_j\}$. Note that an instance of SCP may be defined by a family of subsets $\{M_j\}$ or, alternatively, by a family of subsets $\{N_i\}$.

Suppose the *binary representation* of the SCP solutions is used, that is, genes $g_j \in \{0, 1\}, j \in U$ are the indicators of the elements from U, so that $J(g) = \{j \in U : g_j = 1\}$. If $J(g)$ is a cover, then we assign its fitness $\phi(g) = n - |J(g)|$; otherwise, $\phi(g) = r(g)$, where $r(g) < 0$ is a decreasing function of the number of noncovered elements from M.

Consider a family $\mathcal{B}(n, k)$ of set cover problems introduced by Balas (1984). Here it is assumed that $M = \{1, \ldots, \binom{n}{n-k+1}\}$ and that all $(n - k + 1)$-element subsets of U are given as subsets $N_1, N_2, \ldots, N_{|M|}$. Thus, any collection of less than k elements from U belongs to $U \setminus N_i$ for some $i \in M$ and does not cover the element $i \in M$. At the same time any subset $J \subseteq U$ of size k covers all elements of M and therefore it is an optimal cover. Larger subsets are non-optimal covers.

Since any k-element subset of U is an optimal cover, family $\mathcal{B}(n, k)$ is solvable trivially. Nevertheless this family is known to be hard for general-purpose integer programming algorithms (Balas, 1984; Saiko, 1989). In particular, it was shown in Saiko (1989) that problems from this class are hard to solve using the L-class enumeration method (Kolokolov, 1996). When n is even and $k = n/2$, the L-class enumeration method needs an exponential number of iterations in n. In what follows, we analyze the EA in this special case.

Note that any i-element subset $J \subseteq U$ for $i < k$ leaves $\binom{n-i}{n-(k-1)}$ elements of the ground set uncovered, regardless of the choice of elements in J. So in the case of tournament selection, equivalently to studying the EA on family $\mathcal{B}(n, n/2)$ we may study the EA where the fitness is given by a function of unitation, so that

$$\phi(g) = \begin{cases} R(||g||_1) & \text{if } ||g||_1 \geq n/2; \\ L(||g||_1) & \text{otherwise,} \end{cases}$$

where function R is decreasing, function L is increasing and $L(\frac{n}{2} - 1) < R(n)$.

Consider the point mutation operator with tunable parameter $q > 0$ defined in Subsection 3.1. Let $m = n/2$ and let the thresholds $\phi_0, \phi_1, \ldots, \phi_m$ be equal to fitness of genotypes that contain $0, 1, \ldots, m$ genes "1" accordingly. Note that $J(g)$ is a cover iff $\phi(g) \geq \phi_m$.

We have the following lower bounds: $\alpha_{ij} = 1$ for all $i = 1, \ldots, m, \ j = 0, \ldots, i - 1$; $\alpha_{i,i+1} = (1 - q)(n - i)/n$ for $i = 0, \ldots, m - 1$;

$$\alpha_{ii} = \begin{cases} q + \alpha_{i,i+1} & \text{if } i = 1, \ldots, m - 1; \\ q & \text{if } i = m; \end{cases}$$

$\alpha_{ij} = 0, \ i = 0, \ldots, m - 2, \ j = i + 2, \ldots, m$. These lower bounds α_{ij} coincide with the corresponding cumulative transition probabilities except for level $i = m$, where we pessimistically assume $\alpha_{mm} = q$ (in fact, we could safely put $\alpha_{mm} = 0.5(1 - q) + q$ but $\alpha_{mm} = q$ is chosen to match the model of Ehrenfests in what follows). It is easy to verify that $\mathbf{A}$ satisfies the monotonicity condition when $q \geq 1/(n + 1)$ just as we verified this in the example of monotone mutation in Subsection 3.1.

In case we are interested in runtime bounds for the EA, rather than expected values of vector $\mathbf{z}^{(t)}$, we can assume $\alpha'_{mm} = 1$. All other nonzero lower bounds α_{ij} defined above could be relaxed by putting $\alpha'_{ij} = 1/2$. In this case, we would have the associated Markov chain with a transition matrix $\mathbf{T}'$, the same as in Subsection 5.3, resulting in the same EA runtime bound $O(\lambda n^2)$. We shall avoid these simplifications, however, in order to obtain a tighter runtime bound by means of the following corollary.

COROLLARY 3: *Suppose that the EA with a tournament size $s \geq 1$ uses the point mutation operator with parameter $q \geq 1/(n+1)$. Then given $X^0 = (0, \ldots, 0)$, there exists a constant c, such that the probability to reach an optimum of problem $\mathcal{B}(n, n/2)$ within $\lceil cn \ln n \rceil$ iterations is $\Omega(n^{-0.5})$.*

To prove this corollary, first we will obtain a lower bound on $\mathbf{E}[z_m^{(t)}]$ for $t \to \infty$, using Theorem 2 and the stationary distribution of the associated Markov chain $\mathbf{p}^{(t)} = \mathbf{p}^{(0)} \mathbf{T}^t$. After that, analogously to the proof of Corollary 1, we will compute a lower bound on $\mathbf{E}[z_m^{(t)}]$ for finite t, using Theorem 1.

PROOF OF COROLLARY 3: The Markov chain associated to the set of lower bounds α_{ij} defined above has the following nonzero transition probabilities

$$t_{ii} = q, \quad t_{i,i-1} = (1-q)i/n, \quad t_{i,i+1} = (1-q)(1-i/n), \quad i = 1, \ldots, m-1,$$

$$t_{0,1} = 1-q, \quad t_{m,m-1} = 1-q, \quad t_{mm} = q.$$

All other elements of matrix $\mathbf{T}$ are equal to zero.

The stationary distribution of the associated Markov chain may be found from the well-known model for diffusion of P. Ehrenfest and T. Ehrenfest. Consider n molecules in a rectangular container divided into two equal parts A and B. At any time t, one randomly chosen molecule moves to another part. The state of the system is defined by the number of molecules j, $j = 0, \ldots, n$, in container A. The corresponding random walk has transition probabilities

$$\tau_{j,j-1} = j/n, \quad \tau_{j,j+1} = 1 - j/n, \quad j = 1, \ldots, n-1,$$

$$\tau_{0,1} = 1, \quad \tau_{n,n-1} = 1.$$

The stationary distribution in the model of Ehrenfests (see, e.g., Feller, 1957, Chapter 15, Section 6) is given by $\pi_j := \binom{n}{j}/2^n$, $j = 0, \ldots, n$. Grouping each couple of symmetric states (i.e., the state where A contains j molecules, B contains $n-j$ molecules and the state where A contains $n-j$ molecules and B contains j molecules, $j = 0, \ldots, n/2$) into one state we conclude that the Markov chain with transition matrix $\mathbf{T}$ has the stationary distribution $\mathbf{u} = (2\pi_1, \ldots, 2\pi_m)$ for any $q < 1$. So by Theorem 2, vector $\mathbf{uL}$ is the limiting lower bound for $\mathbf{E}[\mathbf{z}^{(t)}]$ as $t \to \infty$. This allows us to find the vector $\mathbf{v} = \alpha(\mathbf{I} - \mathbf{W})^{-1}$, which is the limit of the right-hand side in Inequality (7) as $t \to \infty$: Recall that the right-hand sides in Inequalities (7) and (8) of Theorems 1 and 2 are equal, given equal matrices $\mathbf{A}$, therefore, $\mathbf{v} = (2\pi_1, \ldots, 2\pi_m) \mathbf{L}$. In particular, $v_m = \frac{\binom{n}{n/2}}{2^{n-1}}$.

We are interested in transient behavior of the EA, so we will obtain a lower bound for the expected population vector $\mathbf{E}[\mathbf{z}^{(t)}]$, given a finite t, using Theorem 1. Consider the matrix norm $||\mathbf{W}||_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^m |w_{ij}|$ which is associated to the vector norm $|| \cdot ||_1$ in the case of left-hand side multiplication of matrices by vectors. For the matrix $\mathbf{W}$, corresponding to the set of lower bounds α_{ij}, defined above, we have $||\mathbf{W}||_\infty = 1 - 2(1-q)/n$, i.e. the condition $\lim_{t\to\infty} ||\mathbf{W}^t||_\infty = 0$ is satisfied for any $q < 1$.

Again let $\mathbf{e} = (1, \ldots, 1)$. By properties of the norms under consideration, $\mathbf{vW}^t \leq ||\mathbf{vW}^t||_1 \mathbf{e} \leq ||\mathbf{v}||_1 \cdot ||\mathbf{W}||_\infty^t \mathbf{e} \leq m||\mathbf{W}||_\infty^t \mathbf{e}$, so by Theorem 1

$$\mathbf{E}[\mathbf{z}^{(t)}] \geq \mathbf{E}[\mathbf{z}^{(0)}]\mathbf{W}^t + \alpha(\mathbf{I} - \mathbf{W})^{-1}(\mathbf{I} - \mathbf{W}^t)$$

$$\geq \alpha(\mathbf{I} - \mathbf{W})^{-1} - \alpha(\mathbf{I} - \mathbf{W})^{-1}\mathbf{W}^t \geq \mathbf{v} - m||\mathbf{W}||_\infty^t \mathbf{e}$$

for any t. With $q = 1/(n+1)$, the average proportion of feasible genotypes is lower-bounded by $v_m - m\left(\frac{n-1}{n+1}\right)^t$ since $||\mathbf{W}||_\infty = 1 - \frac{2(1-q)}{n} = \frac{n-1}{n+1}$. Using the fact that $v_m = \frac{\binom{n}{n/2}}{2^{n-1}}$,

by the Stirling's inequality $\sqrt{2\pi}n^{n+0.5}e^{-n} \leq n! \leq en^{n+0.5}e^{-n}$ we conclude that $v_m = \Omega(n^{-0.5})$. Now assuming that a constant c is so large that $cn \ln n \geq \frac{n+1}{2} \ln \frac{n}{v_m}$, for $t = \lceil cn \ln n \rceil$ we have

$$\frac{v_m}{n} \geq \left(\frac{1}{e}\right)^{\frac{2t}{n+1}} \geq \left(\left(1 - \frac{2}{n+1}\right)^{\frac{n+1}{2}}\right)^{\frac{2t}{n+1}} = \left(\frac{n-1}{n+1}\right)^t,$$

so $\frac{n}{2}\left(\frac{n-1}{n+1}\right)^t \leq \frac{v_m}{2}$ and $\mathbf{E}[z_m^{(t)}] \geq \frac{v_m}{2}$.

By assumption the initial population consists of all-zero strings. Therefore the presence of at least one individual from H_m in the current population implies that an optimal solution to a problem $\mathcal{B}(n, n/2)$ was already found at least once. Thus, in view of Proposition 1, after $\lceil cn \ln n \rceil$ iterations of the EA, the probability of finding an optimum is at least $\Omega(n^{-0.5})$ and the corollary is proved. □

If the EA is restarted with $X^0 = (\mathbf{0}, \ldots, \mathbf{0})$ every $t_{\max} = \lceil cn \ln n \rceil$ iterations, then by Markov inequality the overall runtime of this iterated EA is $O(\lambda n^{1.5} \log n)$ for any λ.

The tools for the non-elitist EA analysis from Corus et al. (2014), Dang and Lehre (2016), and Eremeev (2017) can be adjusted to upper-bound the runtime of the EA on $\mathcal{B}(n, n/2)$, but in such a case, a nonzero selection pressure would be required with a sufficiently large s and the results would hold only for $\lambda = \Omega(\log n)$.

5.5 Upper Bound on Proportion of Optimal Genotypes in Case of ONEMAX

The upper bounds on vector $\mathbf{z}^{(t)}$ obtained in Proposition 4 are not likely to be suitable for obtaining the lower bounds on runtime of the EA in absolute terms due to nonlinearity in the right-hand side of Inequality (13). There are other methods for finding such lower bounds on the runtime proposed (e.g., in Badkobeh et al., 2014, Lehre 2010, and Sudholt, 2013). The upper bounds on vector $\mathbf{z}^{(t)}$, however, may be used for comparison of the EA to the $(1,\lambda)$ EA and the $(1+1)$ EA as it was suggested in Proposition 5.

To illustrate such a comparison, let us consider the EA with bitwise mutation operator Mut in the case of ONEMAX fitness function and assume that $\phi_i := i$, $i = 0, \ldots, n$. Analogously to the notation form Section 3, $P_n^{(\tau)}$ and $Q_n^{(\tau)}$ will stand for the probability to have an optimal current individual on iteration τ of $(1,\lambda)$ EA and on iteration τ of the $(1+1)$ EA, respectively. In these algorithms we assume that the bitwise mutation operator Mut$'$ = Mut is used and the initial solution is chosen uniformly from $\mathcal{X}$. Proposition 5 yields the following:

COROLLARY 4: *Suppose that the fitness function is* ONEMAX *and the initial population of the EA consists of λ copies of the same solution, chosen uniformly from $\mathcal{X}$, and the EA uses the bitwise mutation operator with $p_{\mathrm{m}} = 1/n$. Then for any $t \geq 0$ holds*

$$\mathbf{E}[z_n^{(t+1)}] \leq \frac{1}{e} - \frac{n-2}{e(n-1)}(1 - P_n^{(t)})^s \leq \frac{1}{e} - \frac{n-2}{e(n-1)}(1 - Q_n^{(t\lambda)})^s.$$

In particular, if the tournament size $s = 2$ then $\mathbf{E}[z_n^{(t+1)}] \leq 0.74 P_n^{(t)} + O(n^{-1})$ and $\mathbf{E}[z_n^{(t+1)}] \leq 0.74 Q_n^{(t\lambda)} + O(n^{-1})$.

PROOF: In the case of ONEMAX fitness function the bitwise mutation operator with $p_{\mathrm{m}} = 1/n$ is monotone (Borisovsky and Eremeev, 2008). Application of Proposition 5 yields $\mathbf{E}[z_n^{(t+1)}] \leq \gamma_{nn} - (\gamma_{nn} - \gamma_{n-1,n})(1 - P_n^{(t)})^s$ for the cumulative transition probabilities γ_{ij} associated with this monotone mutation operator. It is easy to see that

$\gamma_{n-1,n} \leq e^{-1}/(n-1)$ and $\gamma_{n,n} \leq e^{-1}$, since $(1-1/n)^n \leq e^{-1}$. Thus, for the $(1,\lambda)$ EA

$$\mathbf{E}[z_n^{(t+1)}] \leq \frac{1}{e} - \left(\frac{1}{e} - \frac{1}{e(n-1)}\right)(1 - P_n^{(t)})^s$$

as required. In the case of $s = 2$ this inequality implies that $\mathbf{E}[z_n^{(t+1)}] \leq \frac{2(n-2)}{e(n-1)}P_n^{(t)} + \frac{1}{e(n-1)} \leq 0.74P_n^{(t)} + O(n^{-1})$. The result for $(1+1)$ EA follows analogously. $\square$

A superiority of the $(1+1)$ EA over other evolutionary algorithms in the case of ONEMAX fitness function and bitwise mutation with $p_m \leq 0.5$ is well-known from Borisovsky (2001), Borisovsky and Eremeev (2008), and Sudholt (2013). Corollary 4 allows to measure the superiority of $(1+1)$ EA and the $(1,\lambda)$ EA over the EA in terms of tail bounds. Note that the tail bounds for the $(1+1)$ EA on ONEMAX are well studied. In particular, the tail bound from Lehre and Witt (2014) implies that there exists such constant $c > 0$ that for any $r \geq 0$ and $\tau < en\ln n - cn - ren$ holds $Q_n^{(\tau)} \leq e^{-r/2}$.

6 Conclusions

In this article, we presented an approximating model of non-elitist mutation-based EA with tournament selection and obtained upper and lower bounds on proportion of sufficiently good genotypes in population using this model. In the special case of monotone mutation operator, the obtained bounds become tight in different situations. The analysis of infinite population EA with monotone mutation suggests an optimal selection mechanism that actually converts the EA into the $(1,\lambda)$ EA.

Applications of the obtained general lower bounds give an exponentially vanishing tail bound for the RLS on unimodal functions and new runtime bounds for the EAs on the 2-SAT problem and on a family of set covering problems proposed by E. Balas.

It is expected that further research will involve applications of the proposed approach to other combinatorial optimization problems, in particular, the problems with regular structure.

Most of the lower and upper bounds on expected proportions of genotypes, obtained in this article, do not take the tournament size into account. It remains an open research question of how to construct the tighter bounds with regards to the tournament size. The subsequent research might benefit from joining the analysis of expectation of population vector with some variance analysis.

It is of interest to compare the tail bounds established in Subsections 5.2 and 5.5 to the tail bounds obtainable using other techniques (e.g., Lehre and Witt, 2014).

Another open question is how to incorporate the crossover operator into the approximating model. For some types of crossover operators, such as those based on solving the optimal recombination problem (Eremeev and Kovalenko, 2014), the lower bounds from this article may be easily extended, ignoring the improving capacity of crossover. It is important, however, to take the positive effect of crossover into account and it is not clear how the monotonicity conditions could be meaningfully extended for this purpose.

Acknowledgments

References

Bäck, T. (1992). The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm. In *Parallel Problem Solving from Nature*, pp. 85–94.

Badkobeh, G., Lehre, P. K., and Sudholt, D. (2014). Unbiased black-box complexity of parallel search. In *Parallel Problem Solving from Nature*, pp. 892–901. Lecture Notes in Computer Science, vol. 8672.

Balas, E. (1984). A sharp bound on the ratio between optimal integer and fractional covers. *Mathematics of Operational Research*, 9(1):1–5.

Borisovsky, P. (2001). Development of evolutionary algorithms and analysis of their precision. Diploma work, Omsk State University. (In Russian.)

Borisovsky, P., and Eremeev, A. (2001). On performance estimates for two evolutionary algorithms. In *European Conference on Evolutionary Computation in Combinatorial Optimization*, pp. 161–171. Lecture Notes in Computer Science, vol. 2037.

Borisovsky, P., and Eremeev, A. (2008). Comparing evolutionary algorithms to the $(1 + 1)$-EA. *Theoretical Computer Science*, 403(1):33–41.

Collatz, L. (1966). *Functional analysis and numerical mathematics*. New York: Academic Press.

Corus, D., Dang, D.-C., Eremeev, A. V., and Lehre, P. K. (2014). Level-based analysis of genetic algorithms and other search processes. In *Parallel Problem Solving from Nature*, pp. 912–921. Lecture Notes in Computer Science, vol. 8672.

Cramer, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Daley, D. J. (1968). Stochastically monotone Markov chains. *Z. Wahrscheinlickeitstheorie und Verw. Gebiete*, 10:307–317.

Dang, D.-C., and Lehre, P. K. (2016). Runtime analysis of non-elitist populations: From classical optimisation to partial information. *Algorithmica*, 75(3):428–461.

Doerr, B., Johannsen, D., and Winzen, C. (2010). Drift analysis and linear functions revisited. *IEEE Congress on Evolutionary Computation*, 1–8.

Eremeev, A. V. (2000). Modeling and analysis of genetic algorithm with tournament selection. In *Artificial Evolution*, pp. 84–95. In Lecture Notes in Computer Science, vol. 1829.

Eremeev, A. (2017). Hitting times of local and global optima in genetic algorithms with very high selection pressure. doi:10.2298/YJOR160318016E.

Eremeev, A., and Kovalenko, J. (2014). Optimal recombination in genetic algorithms for combinatorial optimization problems: Part I. *Yugoslaviam Journal of Operational Research*, 24(1):1–20.

Feller, W. (1957). *An introduction to probability theory and its applications*. New York: John Wiley and Sons.

Garey, M., and Johnson, D. (1979). *Computers and intractability. A guide to the theory of NP-completeness*. New York: W.H. Freeman and Company.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley.

Jansen, T., and Zarges, C. (2012). Fixed budget computations: A different perspective on run time analysis. *Annual Conference on Genetic and Evolutionary Computation*, 1325–1332.

Kolmogorov, A. N., and Fomin, S. V. (1999). *Elements of the theory of functions and functional analysis*. Vol. 2. Mineola, NY: Dover Publications.

Kolokolov, A. A. (1996). Regular partitions and cuts in integer programming. In *Discrete analysis and operations research*, pp. 59–79. Amsterdam: Springer Netherlands.

Lehre, P. K. (2010). Negative drift in populations. In *Parallel Problem Solving from Nature*, pp. 244–253. Lecture Notes in Computer Science, vol. 6238.

Lehre, P. K. (2011). Fitness-levels for non-elitist populations. *Annual Conference on Genetic and Evolutionary Computation*, 2075–2082.

Lehre, P. K., and Witt, C. (2014). Concentrated hitting times of randomized search heuristics with variable drift. In H.-K. Ahn and C.-S. Shin (Eds.), *25th international symposium on algorithms and computation*, pp. 686–697. Cham, Switzerland: Springer.

Moraglio, A., and Sudholt, D. (2015). Principled design and runtime analysis of abstract convex evolutionary search. doi:10.1162/EVCO_a_00169.

Mühlenbein, H., Mahnig, T., and Rodriguez, A. O. (1999). Schemata, distributions and graphical models in evolutionary optimization. *Journal of Heuristics*, 5(2):215–247.

Neumann, F., and Witt, C. (2010). *Bioinspired computation in combinatorial optimization: Algorithms and their computational complexity*. 1st ed. New York: Springer-Verlag.

Nix, A., and Vose, M. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88.

Noschese, S., Pasquini, L., and Reichel, L. (2013). Tridiagonal Toeplitz matrices: Properties and novel applications. *Numerical Linear Algebra with Applications*, 20(2):302–326.

Papadimitriou, C. H. (1991). On selecting a satisfying truth assignment. *32nd Annual Symposium on Foundations of Computer Science*, 163–169.

Rudin, W. (1987). *Real and complex analysis.* 3rd ed. New York: McGraw-Hill.

Rudolph, G. (1994). Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101.

Saiko, L. A. (1989). Studying the cardinality of L-coverings for some covering problems. *Discrete Optimization and Analysis of Complex Systems*, 76–97. (In Russian.)

Schöning, U. (1999). A probabilistic algorithm for $k - SAT$ and constraint satisfaction problems. *40th Annual Symposium on Foundations of Computer Science*, 410–414.

Selman, B., Kautz, H., and Cohen, B. (1996). Local search strategies for satisfiability testing. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 26:521–531. AMS.

Sudholt, D. (2013). A new method for lower bounds on the running time of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 17(3):418–435.

Thierens, D., and Goldberg, D. (1994). Convergence models of genetic algorithm selection schemes. In *Parallel Problem Solving from Nature*, pp. 117–129. Lecture Notes in Computer Science, vol. 866.

Vitányi, P. (2000). A discipline of evolutionary programming. *Theoretical Computer Science*, 241(1–2):3–23.

Vose, M. D. (1995). Modeling simple genetic algorithms. *Evolutionary Computation*, 3(4):453–472.

Wegener, I. (2002). Methods for the analysis of evolutionary algorithms on pseudo-Boolean functions. *Evolutionary Optimization*, 48:349–369.

Witt, C. (2013). Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combinatorics, Probability and Computing*, 22:294–318.

Appendix

In this appendix, we reproduce two results from Borisovsky and Eremeev (2001) and Borisovsky (2001), which are used in Section 3 and a well-known result on eigenvalues of tridiagonal Toeplitz matrices.

The algorithms $(1,\lambda)$ EA and $(1+1)$ EA and probabilities $P_j^{(\tau)}$ and $Q_j^{(\tau)}$, $j = 1, \ldots, m$, $\tau = 0, 1, \ldots$ are defined as in Section 3. For the $(1,\lambda)$ EA and for the $(1+1)$ EA we also define the vectors of probabilities: $\mathbf{P}^{(\tau)} = \left(P_1^{(\tau)}, \ldots, P_m^{(\tau)} \right)$, $\mathbf{Q}^{(\tau)} = \left(Q_1^{(\tau)}, \ldots, Q_m^{(\tau)} \right)$.

The following Theorem 5 from Borisovsky and Eremeev (2001) shows a superiority of the $(1+1)$ EA over the $(1,\lambda)$ EA in the case of monotone mutation operator. For a fair comparison of the algorithms $(1,\lambda)$ EA and $(1+1)$ EA here we allow both of them to make the same number of evaluations of the fitness function, equal to $t\lambda$.

THEOREM 5: *Suppose that the same monotone mutation operator* Mut′ *is used in the $(1+1)$ EA and in the $(1,\lambda)$ EA and* $\mathbf{Q}^{(0)} \geq \mathbf{P}^{(0)}$. *Then* $\mathbf{Q}^{(t\lambda)} \geq \mathbf{P}^{(t)}$ *for any* $t \geq 0$.

The following theorem from Borisovsky (2001) compares the distribution of a fittest individual $g_*^{(t)}$ in the EA population t over Lebesgue subsets compares to such a distribution of the $(1,\lambda)$ EA. Let us define a vector $\mathbf{R}^{(t)}$ for the EA, analogously to vectors $\mathbf{P}^{(t)}$ and $\mathbf{Q}^{(t)}$:

$$\mathbf{R}^{(t)} := \left(\Pr\{g_*^{(t)} \in H_1\}, \ldots, \Pr\{g_*^{(t)} \in H_m\} \right).$$

THEOREM 6: *Suppose that the EA and the $(1,\lambda)$ EA use the same monotone mutation operator* Mut *and* $\mathbf{R}^{(0)} \leq \mathbf{P}^{(0)}$. *Then for any* $t \geq 0$ *holds* $\mathbf{R}^{(t)} \leq \mathbf{P}^{(t)}$, *regardless of selection operator used in the EA.*

The original manuscript (Borisovsky, 2001) is hardly accessible; therefore, we provide the proof of Theorem 6 below.

PROOF: It is sufficient to consider the case of $t = 1$, since the statement for the general case will follow by induction on t. Let $b^{(1,k)}$ denote a genotype with the highest fitness among the first k offspring of $b^{(0)}$ and let $g^{(1,k)}$ be a genotype with the highest fitness among $g_1^{(1)}, \ldots, g_1^{(k)}$ in the EA population X^1, for any $k = 1, \ldots, \lambda$.

a) Let us first assume that $b^{(0)} \in A_i$ and $g_*^{(0)} \in A_i$ for some fixed i and let a genotype g' be chosen by the selection operator of the EA. Then for arbitrary $j = 1 \ldots m$, in view of Proposition 2 we have:

$$\Pr\{\text{Mut}(g') \notin H_j | g_*^{(0)} \in A_i\} \geq \Pr\{\text{Mut}(b^{(0)}) \notin H_j | b^{(0)} \in A_i\}. \tag{23}$$

Note that $\Pr\{g_*^{(1)} \notin H_j | g_*^{(0)} \in A_i\} \geq \Pr\{b^{(1)} \notin H_j | b^{(0)} \in A_i\}$, which may be established by induction on $k = 1, \ldots, \lambda - 1$ using the inequality

$$\Pr\{g^{(1,k+1)} \notin H_j | g_*^{(0)} \in A_i\} = \Pr\{g^{(1,k)} \notin H_j | g_*^{(0)} \in A_i\} \Pr\{\text{Mut}(g') \notin H_j | g_*^{(0)} \in A_i\}$$

$$\geq \Pr\{b^{(1,k)} \notin H_j | b^{(0)} \in A_i\} \Pr\{\text{Mut}(b^{(0)}) \notin H_j | b^{(0)} \in A_i\}$$

$$= \Pr\{b^{(1,k+1)} \notin H_j | b^{(0)} \in A_i\}.$$

b) Let us prove that $\mathbf{P}^{(1)} \geq \mathbf{R}^{(1)}$ for arbitrary initial distributions of the $(1,\lambda)$ EA and the EA, assuming $\mathbf{P}^{(0)} = \mathbf{R}^{(0)}$. We use the total probability formula and the conclusion

of case a):

$$\Pr\{g_*^{(1)} \notin H_j\} = \sum_{i=0}^{m} \Pr\{g_*^{(1)} \notin H_j | g_*^{(0)} \in A_i\} \Pr\{g_*^{(0)} \in A_i\}$$

$$\geq \sum_{i=0}^{m} \Pr\{b^{(1)} \notin H_j | b^{(0)} \in A_i\} \Pr\{b^{(0)} \in A_i\} = \Pr\{b^{(1)} \notin H_j\}. \qquad (24)$$

c) In general, when $\mathbf{P}^{(0)} \geq \mathbf{R}^{(0)}$ let us note that according to Proposition 1 from Borisovsky and Eremeev (2001), in the case of monotone mutation for any $t \geq 1$ we can consider $\mathbf{P}^{(t)}$ as the following function on vector $\mathbf{P}^{(t-1)}$:

$$P_j^{(t)} = 1 - (1 - \gamma_{0j})^\lambda + \sum_{i=1}^{m} ((1 - \gamma_{i-1,j})^\lambda - (1 - \gamma_{ij})^\lambda) P_i^{(t-1)}, \quad j = 1, \ldots, m, \qquad (25)$$

where γ_{ij} are the cumulative transition probabilities of mutation operator Mut. We denote the relationship Eq. (25) by $\mathbf{P}^{(t)} = F(\mathbf{P}^{(t-1)})$ for brevity. Then due to nonnegativity of the multipliers of probabilities $P_1^{(t-1)}, \ldots, P_m^{(t-1)}$ in Eq. (25), we conclude that $\mathbf{P}^{(1)} = F(\mathbf{P}^{(0)}) \geq F(\mathbf{R}^{(0)})$. Finally, note that the result of case b) may be written as $F(\mathbf{R}^{(0)}) \geq \mathbf{R}^{(1)}$, therefore $\mathbf{P}^{(1)} \geq \mathbf{R}^{(1)}$. □

The following result on eigenvalues of tridiagonal Toeplitz matrices may be found, for example, in Noschese et al. (2013).

THEOREM 7: *Suppose an $(n \times n)$-matrix $\mathbf{T}$ is composed of zero elements everywhere except for the diagonal elements, which equal δ, the superdiagonal elements which equal τ and subdiagonal elements which equal σ. Then all of eigenvalues of $\mathbf{T}$ are given by*

$$\lambda_h = \delta + 2\sqrt{\sigma\tau} \, \cos \frac{h\pi}{n+1}, \quad h = 1, \ldots, n.$$