Generalization of the Heavy-Tailed Mutation in the $(1 + (\lambda, \lambda))$ Genetic Algorithm

Anton Eremeev
eremeev@ofim.oscsbras.ru
Novosibirsk State University
Sobolev Institute of Mathematics
Novosibirsk, Russia

Valentin Topchii
topchij@gmail.com
Novosibirsk State University
Sobolev Institute of Mathematics
Novosibirsk, Russia

ABSTRACT

The heavy-tailed mutation operator, proposed by Doerr, Le, Makhmara, and Nguyen (2017) for evolutionary algorithms, is based on the power-law assumption of mutation rate distribution. Here we generalize the power-law assumption on the distribution function of mutation rate. We show that upper bounds on the expected optimization time of the $(1 + (\lambda, \lambda))$ genetic algorithm obtained by Antipov, Buzdalov and Doerr (2022) for the OneMax fitness function do not only hold for power-law distribution of mutation rate, but also for a wider class of distributions, defined in terms of power-law constraints on the cumulative distribution function of mutation rate. In particular, it is shown that, on this function class, the sufficient conditions of Antipov, Buzdalov and Doerr (2022) on the heavy-tailed mutation, ensuring the $O(n)$ optimization time in expectation, may be generalized as well. This optimization time is known to be asymptotically faster than what can be achieved by the $(1 + (\lambda, \lambda))$ genetic algorithm with any static mutation rate.

CCS CONCEPTS

• **Theory of computation → Theory of randomized search heuristics**.

KEYWORDS

Genetic Algorithm, Heavy-Tailed Mutation, Optimization Time

ACM Reference Format:
Anton Eremeev and Valentin Topchii. 2024. Generalization of the Heavy-Tailed Mutation in the $(1 + (\lambda, \lambda))$ Genetic Algorithm. In *Genetic and Evolutionary Computation Conference (GECCO '24 Companion), July 14–18, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3638530.3664095

1 INTRODUCTION

The authors of [1] developed a genetic algorithm $(1+(\lambda, \lambda))$ GA for pseudo-Boolean optimization with a crossover operator, eliminating "unsuccessful" mutations. At each iteration of the $(1 + (\lambda, \lambda))$ GA, λ offspring are generated from a single parent individual independently of each other, at the Hamming distance ℓ from the parent. It is assumed that $\ell \sim \mathrm{Bin}(n, p)$, n is the length of the bit-string and p is the mutation parameter. Next, the best individual in terms of fitness is selected from these solutions and the crossover operator is applied to it, given a crossover parameter c. With probability c the crossover operator uses bits from the best child, and with probability $1-c$, it uses bits from the parent solution. This way λ individuals are created, and the best of the λ individuals is accepted as a new parent, provided it is at least as fit as the previous parent.

In the present work, we consider the genetic algorithm $(1 + (\lambda, \lambda))$ GA from [1], combined with the fast mutation operator. In [2] for $(1 + (\lambda, \lambda))$ GA with fast mutation and a special random choice of values λ and p, obtained an upper bound for the optimization time of an order $O(n)$ in the case of OneMax. This is less than the optimization time of $(1 + (\lambda, \lambda))$ GA with any fixed probability of mutation. In this algorithm, both the population size λ and the parameter of fast mutation p have a truncated power-law distribution with upper bounds $\lambda \leq u_n$ and $p \leq u_n/n$ respectively. The linear upper bound in [2] holds when the power exponent β satisfies the inequalities $2 < \beta < 3$ and $u_n \geq \ln^{1/(3-\beta)} n$.

The main result of this work (Theorem 2) shows that upper bounds for optimization time $(1 + (\lambda, \lambda))$ GA, similar to those obtained in [2], are valid not only in the case of truncated power-law distributions of the random variables λ and p, but also for a wider class of distributions described in terms power-law constraints on the cumulative distribution function. The linear upper bound we obtained for the optimization time, similarly to the linear bound from [2], is asymptotically smaller than the optimization time of $(1 + (\lambda, \lambda))$ GA for any static mutation parameter p.

Let us denote $\mathbb{N}_m := \{k : k \in \mathbb{N}, k \leq m\}$, $S := \{0, 1\}^n$, $|S| = 2^n$. Hamming norm and Hamming distance are $|x| := \sum_{i=1}^n x_i$ and $|x - y| := \sum_{i=1}^n |x_i - y_i|$ for $x, y \in S$. We denote a solution with ones in all bits as x^*, and define $Z_s := \{x \in S : |x - x^*| = s\}$, which is the set of solutions with exactly s zeros, $s = 0, \ldots, n$.

In what follows, we consider and evaluate the characteristics of the algorithm optimizing a fitness function f from the OneMax family, $f(x(i)) = |x(i)|$. Let $\lambda(n)$, $n \in \mathbb{N}$, be a set collectively independent random variables (hereinafter, random variables are referred to as r.v.) with $\mathbf{P}(\lambda(n) = k) = p_{n,k}$, where $u_n \leq 0.5n$, $p_{n,k} = 0$ for $k \in \mathbb{N} \setminus \mathbb{N}_{u_n}$ and $p_{n,u_n} > 0$. The algorithm under consideration coincides with Algorithm 1 from [2], but we depart from explicit power expressions for $p_{n,k}$, instead we use only constraints on the cumulative distribution function of $\lambda(n)$.

Positive constants independent of n we will denote as C or C_i.

Definition 1. *For r.v. $\lambda(n)$ we define the following conditions:*
- $\mathcal{A}_2^c$ *if* $\exists\, C > 0$ *is such that* $\mathrm{E}\lambda^2(n) < C$.
- $\mathcal{A}_2^{\mathrm{pow}}$, *if* $\exists\, C_1(\beta) > 0, C_2(\beta) > 0$ *and* $b_0 > 0$ *such that for all*

GECCO '24 Companion, July 14–18, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0495-6/24/07
https://doi.org/10.1145/3638530.3664095

sufficiently large $n \in \mathbb{N}$ and $b \in [b_0, u_n]$ we have

$$\mathbf{E}\lambda^2(n) \geq C_1(\beta)u_n^{3-\beta}, \;\; \mathbf{P}(b/2 \leq \lambda(n) \leq b) \geq C_2(\beta)b^{1-\beta},$$

where $\beta \in (1,3)$, $u_n \leq n/2$, and $u_n \to \infty$ as $n \to \infty$.

We generalize Theorems 5 and 6 [2], where $u_n \to \infty$ as $n \to \infty$,

$$\mathbf{P}(\lambda(n) = k) = p_{n,k} = C_{\beta,u_n} k^{-\beta}, k \in \mathbb{N}_{u_n}, \tag{1}$$

to wider classes of distributions of series of the r.v. $\lambda(n)$, $n \in \mathbb{N}$.

Condition $\mathcal{A}_2^c$ means that $\mathbf{E}\lambda^2(n)$ is uniformly bounded and there are no other restrictions on $p_{n,k}$, which holds for $\beta > 3$ in (1).

Condition $\mathcal{A}_2^{\text{pow}}$ implies that the second moment of the r.v. $\lambda(n)$ is unbounded as n grows, and $0 < 3 - \beta < 2$. It imposes restrictions on the cumulative distribution function for r.v. $\lambda(n)$ without fixation of specific values for the probabilities $p_{n,k}$, some of which may even equal to zero. Conditions (1) from [2] are more restrictive, since they have the form $p_{n,k} = C(\beta, u_n)k^{-\beta}$, $k \in \mathbb{N}_{u_n}$, and $C(\beta, u_n)$ is asymptotically constant for $n \to \infty$.

Let algorithm $\mathcal{A}$ denote the fast $(1 + (\lambda, \lambda))$ GA from [2], where instead of the truncated power-law distribution we use one of the distributions satisfying the power-law constraints on the cumulative distribution function from Definition 1.

2 THE MAIN RESULTS

Let $\ell_{\lambda(n)}(s)$ denote the number of iterations of Algorithm 1 from [1] starting from an individual $x \in Z_s$ till the first improvement of the fitness function for fixed n and $\lambda(n)$, and let us denote the probability of improvement at the first iteration by $p_{\lambda(n)}(s) = \mathbf{P}(\ell_{\lambda(n)}(s) = 1)$. In the case of OneMax, the probabilities $p_{\lambda(n)}(s)$ are the same for all $x \in Z_s$. Note that with a random choice of $\lambda(n)$, the r.v. $\ell_{\lambda(n)}(s)$ is not as important as the event $\{\ell_{\lambda(n)}(s) = 1\}$.

By Lemma 7 from [1] with a fixed (not random) value $\lambda = \lambda(n)$ for any n, with mutation parameter $p = \lambda(n)/n$ and crossover parameter $c = 1/\lambda(n)$ and by Lemma 2 from [2] we obtain

$$p_{\lambda(n)}(s) \quad \geq \quad C_1\lambda^2(n)s/n, \text{ at } \lambda^2(n)s/n < 1, \tag{2}$$

$$p_{\lambda(n)}(s) \quad \geq \quad C_2, \text{ with } \lambda^2(n)s/n \geq 1. \tag{3}$$

For the r.v. $\lambda(n)$, we denote the average probability of fitness improvement in one iteration as $p_n(s) := \mathbf{E}_{\lambda(n)}p_{\lambda(n)}(s)$. The sequence of iterations until the first improvement of the fitness function is a sequence of independent trials up to the first success, with success probabilities $p_n(s)$. The number of elements $\ell_n(s)$ in this sequence is geometrically distributed, $\mathbf{P}(\ell_n(s) = 1) = p_n(s)$ and

$$\mathbf{E}\ell_n(s) = (1 - p_n(s))p_n^{-1}(s) + 1 = p_n^{-1}(s). \tag{4}$$

Given some fixed $m_0, m_1 \in \mathbb{N}$, let there exist $C_3 > 0$ such that for all $n \in \mathbb{N} \setminus \mathbb{N}_{m_1}$ the following inequalities hold

$$\mathbf{P}(\lambda(n) \in \mathbb{N}_{m_0}) \geq C_3. \tag{5}$$

Markov inequality implies (5), if $\mathbf{E}\lambda(n) \leq C_4$, $\forall n \in \mathbb{N}$ for $C_4 < \infty$.

If condition (5) is satisfied, then taking into account the relation (4) for all $n \in \mathbb{N} \setminus \mathbb{N}_{m_1}$ and $s \in \mathbb{N}_n$ the following inequalities hold

$$p_n(s) \geq \mathbf{E}_{\lambda(n)}\{p_{\lambda(n)}(s); \lambda(n) \in \mathbb{N}_{m_0}\} \geq \frac{C_5 s}{n}, \; \mathbf{E}\ell_n(s) \leq \frac{n}{C_6 s}. \tag{6}$$

Let $\tau_i(n)$ be the number of iterations until the first hitting x^* if the process starts from an individual $x^{(0)} \in Z_i$. According to the total probability formula for $\tau(n)$, which is the number of iterations until the first hitting x^*, starting from a random individual $x^{(0)}$, we have

$$\mathbf{E}\tau(n) = 2^{-n}\sum_{i=0}^{n} C_n^i \mathbf{E}\tau_i(n) \leq 2^{-n}\sum_{i=0}^{n} C_n^i \sum_{s=0}^{i} \mathbf{E}\ell_n(s)$$

$$= 2^{-n}\left(\sum_{s=1}^{n\epsilon} + \sum_{s=n\epsilon+1}^{n}\right)\mathbf{E}\ell_n(s)\sum_{i=s}^{n} C_n^i \leq \sum_{s=1}^{n\epsilon}\mathbf{E}\ell_n(s) + C_5 n, \tag{7}$$

where $\epsilon \in (0,1)$ is an arbitrary fixed value.

Lemma 1. *If $\lambda(n)$ meets condition $\mathcal{A}_2^{\text{pow}}$ in the Algorithm $\mathcal{A}$, then*

$$\mathbf{E}\ell_n(s) \leq \frac{C_1^* n}{u_n^{3-\beta}\cdot s}, \quad \text{if } u_n^2 s/n < 1, \tag{8}$$

$$\mathbf{E}\ell_n(s) \leq \frac{C_2^* n}{(n/s)^{(3-\beta)/2}\cdot s}, \quad \text{if } u_n^2 s/n > 1. \tag{9}$$

The theorem below follows from (6), (7), (8) and (9).

Theorem 1. *The expected number of iterations till finding the optimum of* OneMax *by Algorithm $\mathcal{A}$ satisfies the upper bounds*

$$\mathbf{E}\tau(n) \quad = \quad O(n \ln n), \text{ subject to } \mathcal{A}_2^c; \tag{10}$$

$$\mathbf{E}\tau(n) \quad = \quad O(n) + O\left(\frac{n \ln n}{u_n^{3-\beta}}\right), \text{ subject to } \mathcal{A}_2^{\text{pow}}. \tag{11}$$

Bound (11) differs from the analogous bound in Theorem 5 [2] by the second term of (11). Theorem 5 [2] may be considered as a special case of (11) when $\mathbf{E}\lambda(n) < \infty$ and $u_n \geq \ln^{1/(3-\beta)} n$, which implies that $u_n^{\beta-3}\ln n = O(1)$ and the second term turns into $O(n)$.

Let $T(n)$ be the number of fitness evaluations before visiting $x^* \in \{0,1\}^n$, and $T^{op}(n)$ is the number of operations till this event.

If the number of fitness evaluations on each iteration does not exceed $2\lambda(n)$, then the number of operations depends on the implementation of the computational algorithm. Let us assume that there exists a function $\phi(\lambda(n), n)$ such that the number of operations per iteration does not exceed $\phi(\lambda(n), n)$. Denoting $\phi_n := \mathbf{E}_{\lambda(n)}\phi(\lambda(n), n)$, we can prove that for some constant C^*:

$$\mathbf{E}T(n) \leq C^*\mathbf{E}\tau(n)\mathbf{E}\lambda(n), \;\; \mathbf{E}T^{op}(n) \leq C^*\mathbf{E}\tau(n)\phi_n. \tag{12}$$

This results in a generalization of Theorem 6 from [2].

Theorem 2. *Bounds (12) on the average number of fitness evaluations $T(n)$ and the average number of operations $T^{op}(n)$ are valid for Algorithm $\mathcal{A}$, optimizing* OneMax *fitness function.*

Our result, e.g. is valid in the case of $p_{n,k} = C^{(*)}(\beta, u_n)k^{1-\beta}$ for $k \in \mathbb{N}_{u_n}^{(3)}$ and $u_n \in \mathbb{N}^{(3)}$, where in the accepted notation the set of natural numbers is replaced by $\mathbb{N}^{(3)} = \{3^\ell, \ell = 0, 1, 2, \cdots\}$.

ACKNOWLEDGMENTS

The work is supported by the Mathematical Center in Akademgorodok under the agreement No. 075-15-2022-282 with the MSHE RF.

REFERENCES

[1] B. Doerr, C. Doerr, and F. Ebel. *From blackbox complexity to designing new genetic algorithms.* Theoretical Computer Science, (2015) 567, 87-104.
[2] D. Antipov, M. Buzdalov, B. Doerr, *Fast mutation in crossover-based algorithms*, Algorithmica, **84**:6 (2022), 1724–1761.