UPPER BOUNDS ON EXPECTED HITTING TIME OF TARGET SUBSETS BY GENETIC ALGORITHM

D.-C. Dang, A.V. Eremeev, P.K. Lehre

Keywords: combinatorial optimization problem, levels partition, drift analysis.

The paper is devoted to upper bounds on run-time of Non-Elitist Genetic Algorithms until some target subset of solutions is visited for the first time. By means of drift analysis, the previously known upper bounds are improved and extended.

Introduction. The genetic algorithm (GA) proposed by J. Holland [5] is a randomized heuristic search method, based on analogy with the genetic mechanisms observed in nature and employing a population of tentative solutions. Different modifications of GA are widely used in areas of operations research, pattern recognition, artificial intelligence etc.

In this paper, the genetic algorithms are studied on a wide class of combinatorial optimization problems w.r.t. expected number of tentative solutions visited until a feasible solution with required properties is found for the first time. The target sets of solutions may be defined by an approximation ratio, by local or global optimality conditions, etc. The main result is obtained by combining the approaches proposed in [2,6], and applies to a wider classes of algorithms and problems compared to those addressed in [2,6].

Combinatorial Optimization Problems. In this paper, the combinatorial optimization problems are viewed in the following formulation:

$$\max\{F(x) : x \in \mathrm{Sol}\}, \tag{1}$$

where $\mathrm{Sol} \subseteq \mathcal{X}$ is the set of feasible solutions, $\mathcal{X} = \{0,1\}^n$ is the search space, $F(\cdot)$ is the objective function. Without loss of generality we will consider only the maximization problems. The results will hold for the minimization problems as well.

Let a neighborhood $\mathcal{N}(y) \subseteq \mathrm{Sol}$ be defined for every $y \in \mathrm{Sol}$. The mapping $\mathcal{N} : \mathrm{Sol} \to 2^{\mathrm{Sol}}$ is called the *neighborhood mapping*. If the inequality $F(y) \leq F(x)$ holds for all neighbors $y \in \mathcal{N}(x)$ of a solution $x \in \mathrm{Sol}$, then x is called a local optimum w.r.t. $\mathcal{N}$.

Genetic Algorithms. The GA operates with populations $P^t = (x^{1,t}, \ldots, x^{\lambda,t})$, $t = 0, 1, \ldots$, which consist of λ *genotypes*. In terms of the present paper the genotypes are elements of $\mathcal{X}$. For convenience we assume that the number of genotypes λ is even.

In a *selection* operator $\mathrm{Sel} : \mathcal{X}^\lambda \to \{1, \ldots, \lambda\}$, each parent is independently drawn from the previous population P^t where each individual in P^t is assigned a selection probability depending on its *fitness* $f(x)$. Usually a higher fitness value of an individual implies higher (or equal) selection probability. We assume that the fitness function is defined on the basis of objective function in such a way that if $x \in \mathrm{Sol}$ then $f(x) = F(x)$; otherwise (i.e. if $x \notin \mathrm{Sol}$), the fitness incorporates some penalty, which ensures that $f(x) < \min_{y \in \mathrm{Sol}} F(y)$.

A pair of offspring genotypes is created using the randomized operators of crossover $\mathrm{Cross} : \mathcal{X} \times \mathcal{X} \to \mathcal{X} \times \mathcal{X}$ and mutation $\mathrm{Mut} : \mathcal{X} \to \mathcal{X}$. In general, we assume $\mathrm{Cross}(x, y)$ and $\mathrm{Mut}(x)$ to be efficiently computable randomized routines.

When a population $P^t = (x^{1t}, \ldots, x^{\lambda t})$ is constructed, the GA proceeds to the next iteration $t + 1$. An initial population P^0 is generated randomly. One of the ways of initialization consists, e.g. in independent choice of all bits in genotypes.

In theoretical analysis of the $\mathcal{G}A$ it is often assumed that the termination condition is never met. The termination condition, however, may be required to stop a genetic algorithm when a solution of sufficient quality is obtained or the computing time is limited, or because the population is "trapped" in some unpromising area and it is preferable to restart the search (see e.g. [7]). In order to make multiple independent runs of the $\mathcal{G}A$, we will consider the Iterated Non-Elitist $\mathcal{G}A$ with the following outline.

Algorithm Iterated $\mathcal{G}A$

Repeat:
Generate the initial population P^0, assign $t := 1$.
While termination condition $t > t_{\max}$ is not met **do:**
Iteration t.
For j from 1 to $\lambda/2$ do:
Selection: $i := \mathrm{Sel}(P^{t-1})$, $i' := \mathrm{Sel}(P^{t-1})$.
Crossover: $(x, y) := \mathrm{Cross}(x^{i,t-1}, x^{i',t-1})$.
Mutation: $x^{2j-1,t} := \mathrm{Mut}(x)$, $x^{2j,t} := \mathrm{Mut}(y)$.
End for.
$t := t + 1$.
End while.
Until false.

In what follows, $\mathcal{G}A$ denotes the above algorithm with $t_{\max} = \infty$, i.e. the termination condition is never used. As the examples of crossover and mutation we consider the well-known operators of bitwise mutation Mut^* and the single-point crossover Cross^* from [4].

The single-point crossover computes $(x', y') = \mathrm{Cross}^*(x, y)$, given $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$, so that with a given probability p_{c},

$$x' = (x_1, ..., x_\chi, y_{\chi+1}, ..., y_n), \quad y' = (y_1, ..., y_\chi, x_{\chi+1}, ..., x_n),$$

where the random number χ is chosen uniformly from 1 to $n-1$. With probability $1 - p_{\mathrm{c}}$ both parent individuals are copied without any changes, i.e. $x' = x,\ y' = y$.

The bitwise mutation operator Mut^* computes a genotype $y = \mathrm{Mut}^*(x)$, where independently of other bits, each bit $y_i,\ i \in [n]$, is assigned a value $1 - x_i$ with probability p_{m} and with probability $1 - p_{\mathrm{m}}$ it keeps the value x_i. Here and below we use the notation $[n] := \{1, 2, ..., n\}$ for any positive integer n. Tunable parameter p_{m} is also called *mutation rate*.

The following condition holds for many well-known crossover operators: there exists a positive constant ε_0 which does not depend on problem instance, such that the output of crossover $(x', y') = \mathrm{Cross}(x, y)$ satisfies the inequality

$$\Pr\Big(\max\{f(x'), f(y')\} \geq \max\{f(x), f(y)\}\Big) \geq \varepsilon_0 \tag{2}$$

for any $x, y \in \mathcal{X}$.

Condition (2) suggests that the fitness of at least one of the genotypes resulting from crossover $(x', y') = \mathrm{Cross}(x, y)$ is not less than the fitness of the parents $x, y \in \mathcal{X}$ with probability at least ε_0. This condition is fulfilled for the single-point crossover with $\varepsilon_0 = 1 - p_{\mathrm{c}}$, if $p_{\mathrm{c}} < 1$ is a constant. Condition (2) is also satisfied with $\varepsilon_0 = 1$ for the optimized crossover operators, where at least one of the two offspring is computed as a solution to optimal recombination problem (see e.g., [3]).

Drift Analysis of Non-Elitist Genetic Algorithm. This section generalizes the drift analysis of mutation-based Non-Elitist Genetic Algorithm proposed in [6] adapting it to the two-offspring crossover and making it applicable to Iterated $\mathcal{G}A$, which allows us to deal with both feasible and infeasible solutions.

Suppose that for some m there is an ordered partition of $\mathcal{X}$ into subsets $(A_0, \ldots, A_{m+1})$ called *levels*. Level A_{m+1} will be the target level in subsequent analysis. The target level may be chosen as the set of solutions with maximal fitness or the set of local optima or the set of ρ-approximation solutions for some approximation factor $\rho > 1$. A well-known example of partition is the *canonical* partition, where each level regroups solutions having the same fitness value (see e.g. [6]). For $j \in [m+1]$ we denote by $H_j := \cup_{i=j}^{m+1} A_i$, the union of all levels starting from level j.

Given a levels partition, there always exists a total order "$\succ$" on $\mathcal{X}$, which is aligned with $(A_0, \ldots, A_{m+1})$ in the sense that $x \succ y$ for any $x \in A_j$, $y \in A_{j-1}$, $j \in [m+1]$. W.l.o.g. in what follows the elements in a population vector $P^t \in \mathcal{X}^\lambda$ will be assumed to form a non-increasing sequence $x^{1t}, x^{2t}, \ldots, x^{\lambda t}$ in terms of "$\succ$" order: $x^{1t} \succeq x^{2t} \succeq \ldots \succeq x^{\lambda t}$. For any constant $\gamma \in (0,1)$, the individual $x^{\lceil \gamma\lambda \rceil, t}$ will be referred to as the γ-ranked individual in population P^t.

For any $\gamma \in (0,1)$ and population P of size λ, the *selective pressure* $\beta(\gamma, P)$ of a selection operator Sel is defined as the probability of selecting an individual from P that belongs to the same or higher level as the individual with rank $\lceil \gamma\lambda \rceil$.

Theorem 1 *Given a partition $(A_0, \ldots, A_{m+1})$ of $\mathcal{X}$, let there exist parameters $s_1, \ldots, s_m$, $s_*, p_0, p_1, \varepsilon \in (0,1]$, and two positive constants γ and δ such that*

(C1) $\Pr(Mut(x) \in H_{j+1}) \geq s_j \geq s_*$ *for any* $x \in A_j$, $j \in [m]$,

(C2) $\Pr(Mut(x) \in H_j) \geq p_0$ *for any* $x \in A_j$, $j \in [m]$,

(C3) $\beta(\gamma', P) \geq 2\frac{1+\delta}{\varepsilon p_0}\gamma'$ *for any* $\gamma' \in (0, \gamma]$ *and* $P \in (\mathcal{X} \backslash A_{m+1})^\lambda$,

(C4) $\lambda \geq \frac{4(1+\delta)}{\gamma\delta^2} \max\left\{\ln\left(\frac{m}{s_*}\right), \frac{2^6(2+\delta)(1+\delta)}{\gamma\delta^3}\right\}$,

(C5) $\Pr\Big(Cross(x,y) \in (H_j \times \mathcal{X}) \cup (\mathcal{X} \times H_j)\Big) \geq \varepsilon$, $(x,y) \in (H_j \times \mathcal{X}) \cup (\mathcal{X} \times H_j)$, $j \in [m]$,

(C6) $\Pr\left(x^{\lceil \gamma\lambda \rceil, 0} \in H_1\right) \geq p_1$,

then, assuming $t^* := \frac{2+\delta}{\delta/2}\left(m(\lambda+1) + \sum_{j=1}^{m} \frac{p_0}{\lambda s_j (1+\delta)}\right)$, the probability that the $\mathcal{G}A$ population contains an individual from A_{m+1} in one of the first t^* generations is at least $p_1/4$.

Condition (C1) requires that for each level j, there is a lower limit s_j on the "upgrade probability" from level j. Condition (C2) requires that there exists a lower limit p_0 on the probability that the individual will not "downgrade" to a lower level. Condition (C3) requires that the selective pressure induced by the selection mechanism is sufficiently strong. Condition (C4) requires that the population size λ is sufficiently large. Condition (C5) is a level-based analog of inequality (2) and follows from (2) with $\varepsilon = \varepsilon_0$ in the case of the canonical partition. Condition (C6) ensures that at least a fraction γ of initial population is "above" level 0 with probability not less than p_1. The level subset A_0 here may be used to encompass the set of infeasible solutions.

Corollary 1 *Suppose that Conditions (C1)-(C6) of Theorem 1 hold. Let T be the number of function evaluations until the subset A_{m+1} is reached for the first time in the Iterated $\mathcal{GA}$ with $t_{\max} = t^*$. Then $E[T] \leq c\left(m\lambda^2 + \sum_{j=1}^{m} s_j^{-1}\right)$ for some constant $c > 0$.*

In this paper, we consider the tournament selection, (μ, λ)-selection and exponential ranking selection. In *k-tournament selection*, k individuals are sampled uniformly at random with replacement from the population, and the fittest of these individuals is returned. In *(μ, λ)-selection*, parents are sampled uniformly at random among the fittest μ individuals in the population. A real-valued function α on $[0, 1]$ is called a ranking function if $\alpha(x) \geq 0$ for all $x \in [0, 1]$, and $\int_0^1 \alpha(x)dx = 1$. In ranking selection with ranking function α, the probability of selecting individuals ranked γ or better is $\int_0^\gamma \alpha(x)dx$. We define *exponential ranking* parameterised by $\eta > 0$ as $\alpha(\gamma) := \eta e^{\eta(1-\gamma)}/(e^\eta - 1)$.

Suppose a maximization problem (1) is given and a neighborhood mapping $\mathcal{N}$ is defined. Let s be a lower bound on the probability that a mutation operator transforms a given feasible solution x into a specific neighbor $y \in \mathcal{N}(x)$, i.e.

$$s \leq \min_{x \in \mathrm{Sol},\ y \in \mathcal{N}(x)} \Pr(\mathrm{Mut}(x) = y). \tag{3}$$

The greater the value s, the more consistent is the mutation with the mapping $\mathcal{N}$.

In many well-known combinatorial optimization problems, such as the Maximum Satisfiability Problem, the Maximum Cut Problem and the Ising Spin Glass Model, the set of feasible solutions is the whole search space, i.e. $\mathrm{Sol} = \mathcal{X}$.

Corollary 2 *Let $\mathrm{Sol} = \mathcal{X}$, Conditions (C5) and (3) be satisfied, $\Pr(Mut(x) = x) \geq p_0$ for any $x \in \mathcal{X}$ for some constant p_0, and assume that the $\mathcal{GA}$ is using either k-tournament selection with $k \geq \frac{2(1+\delta')}{\varepsilon p_0}$, or (μ, λ)-selection with $\frac{\lambda}{\mu} \geq \frac{2(1+\delta')}{\varepsilon p_0}$ or exponential ranking selection with $\eta \geq \frac{2(1+\delta')}{\varepsilon p_0}$ for some constant $\delta' > 0$. Then there exist constants c and c', such that for population size $\lambda \geq c' \ln\left(\frac{m}{s}\right)$, a local optimum is reached for the first time after at most $c(m\lambda^2 + \frac{m}{s})$ fitness evaluations in expectation.*

A similar result for the $\mathcal{GA}$ with tournament selection was obtained in [2] without use of drift analysis. In particular, Lemma 2.1 in [2] implies that with appropriate settings of parameters, the $\mathcal{GA}$ reaches a local optimum for the first time within $O\left(\frac{m \ln m}{s}\right)$ fitness evaluations on average. The bound from Corollary 2 in the present paper grows slower if s is at least linear in m. (Note that the size of the well-known neighborhoods usually grows as some polynomial of m.) The run-time bound for the $\mathcal{GA}$ with the single-offspring crossover may be further tightened using more refined drift analysis proposed in [1].

Now let us turn to the general case where Sol may be a proper subset of $\mathcal{X}$.

Corollary 3 *Let the constant $\varepsilon_0 > 0$ and the lower bound s satisfy (2) and (3) and assume that the Iterated $\mathcal{GA}$ is using either k-tournament selection with $k \geq \frac{2(1+\delta')}{\varepsilon_0 s}$, or (μ, λ)-selection with $\frac{\lambda}{\mu} \geq \frac{2(1+\delta')}{\varepsilon_0 s}$ or exponential ranking selection with $\eta \geq \frac{2(1+\delta')}{\varepsilon_0 s}$ for some constant $\delta' > 0$. Suppose that with probability $p_1 > 0$ at least a constant fraction of individuals in initial population P^0 are feasible, and $\Pr(Mut(x) = x) \geq p_0 > 0$ for any $x \in \mathcal{X}$ for some positive constant p_0.*

Then there exist three constants c, b' and c' such that for population size $\lambda \geq c' \ln\left(\frac{m}{s}\right)$ and $t_{\max} = b'\left(m\lambda + \frac{m}{\lambda s}\right)$, a local optimum is reached for the first time after at most $cp_1\left(m\lambda^2 + \frac{m}{s}\right)$ fitness evaluations in expectation.

Conclusions. The obtained results imply that if a problem is polynomially bounded and feasible solutions are easy to find, then a local optimum is computable in expected polynomial time by the Iterated GA with tournament selection or (μ, λ)-selection or exponential ranking selection. In particular, given suitable parameters, these versions of Iterated GA find the solutions with constant approximation ratio to any problem from the class of combinatorial optimization problems with guaranteed local optima (GLO) in expected polynomial time.

This research received funding from the European Union 7-th Framework Program (FP7/2007-2013) grant agreement 618091 (SAGE), from EPSRC grant EP/F033214/1 (LANCS), and from Russian Foundation for Basic Research grant 15-01-00785.

BIBLIOGRAPHY

1. Corus D., Dang D.-C., Eremeev A.V., Lehre P.K. Level-based analysis of genetic algorithms and other search processes. Proc. of Parallel Problem Solving from Nature (PPSN XIII). Springer, LNCS Vol. 8672, 2014. – P. 912-921.
2. Eremeev A.V. Analysis of Evolutionary Methods of Solution for Combinatorial Optimization Problems. Dr. Sci. Dissertation, speciality 05.13.17, 2013, Omsk, 300 p. (in Russian)
3. Eremeev A.V., Kovalenko J.V. Optimal recombination in genetic algorithms for combinatorial optimization problems: Part I // Yugoslav Journal of Operations Research. – 2014. – Vol. 24, N 1. – P. 1-20.
4. Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. – Reading, MA: Addison-Wesley, 1989.
5. Holland J. Adaptation in Natural and Artificial Systems. – Ann Arbor: University of Michigan Press, 1975.
6. Lehre P.K. Fitness-levels for non-elitist populations. In: Proceedings of 13th Conf. on Genetic and Evolutionary Computation (GECCO), 2011. – P. 2075–2082.
7. Vitányi P. M. B.: A discipline of evolutionary programming // Theor. Comp. Sci. – 2000. – Vol. 241, N 1–2. – P. 3–23.

Dang Duc-Cuong, Lehre Per Kristian,
School of Computer Science, University of Nottingham,
Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, UK,
phone: +44 0115 84 68376,
E-mail: pszpl@nottingham.ac.uk, pszdd@nottingham.ac.uk

Eremeev Anton Valentinovich,
Sobolev Institute of Mathematics, Omsk Branch,
13, Pevtsov str., Omsk, 644099, Russia,
phone: +7 (3812) 236739, fax: +7 (3812) 234584.
E-mail: eremeev@ofim.oscsbras.ru