Asymptotic Correctness of the Evolutionary Algorithm in Banach Space

Jarosław Arabas

Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland,
email: jarabas@ise.pw.edu.pl

Abstract. The paper gives two theorems for the asymptotic correctness of an evolutionary algorithm (EA) that processes chromosomes from an arbitrary Banach space. It is argued that even if the mutation cannot yield an arbitrarily far offspring, the EA may be asymptotically correct provided that the selection is nonelitist and each feasible individual may reproduce with nonzero probability. An illustrative example accompanies the paper.

1 Introduction

When considering a stochastic optimization method one is interested in its asymptotic correctness [5, 8]. This property means that the optimization method will generate at least one point in each arbitrary subset of the search space when the iteration number goes to infinity. When a method is asymptotically correct, the probability that it will generate at least one point arbitrarily close to the global optimum approaches one, as time increases.

This paper is devoted to asymptotic correctness of an evolutionary algorithm (EA) and its implication for the convergence analysis. Convergence of the EA was extensively discussed by many authors. Most of the work was focused on EAs of a special type, and different mathematical approaches were used for proving this property. Characterization of all approaches to modeling EA behavior falls far beyond the scope of this paper, therefore we will name only few of the most widely known.

In the field of genetic algorithms (GA) with binary chromosomes, schema theorem [4, 9] was the first attempt to explain their behavior. According to this theorem, sets of chromosomes (defined by similarity patterns called schemata) with higher average fitness value are allocated higher expected number of copies in the next generation. Thus, through subsequent generations, a tendency is expected for the percentage of schemata with higher average fitness to increase. Another approach to modeling the GA with binary encoding was based on the theory of the Markov chains [8, 10]. The GA is modeled as a random process in a specific space of states. Each state is identified by the vector of bits created by setting all chromosomes in the population one after another. It is proved that under certain conditions there exist absorbing states of such random process and each absorbing state represents a population containing the chromosome with the best fitness.

In the field of evolution strategies, convergence analysis is usually based on direct computation of the probability of generating a point in the neighborhood of the global maximum. Perhaps the most widely known result is that the best point in the population of an elitist $(\mu+\lambda)$ strategy will converge almost surely to the neighborhood of the global maximum provided that the mutation distribution is positive defined (e.g. Gaussian) [7, 8].

The theoretical results mentioned above considered special types of EA, where chromosomes are either vectors of bits or real numbers. On the other hand, a variety of chromosome encoding methods has been proposed, and it seems that the ability to process solutions represented in a nonstandard, usually problem-specific way, is one of the main advantages of EA [2, 6]. Therefore there is a need for a theory that will explain the behavior of an EA in a more general way in order to understand such nonstandard cases.

In this paper a measurable Banach space is considered. Such space is sufficiently general to include all chromosome representations that maintain a constant set of parameters. The paper gives and discusses two alternative theorems of the EA asymptotic convergence, based on different assumptions. In the first theorem, mutation is assumed to have its distribution positively defined in the whole search area, and no restriction on the selection is made. The second theorem makes weaker assumptions about the mutation distribution, but the nonelitist soft selection (i.e. nonzero probability of reproducing any individual) is required instead.

Organization of the paper is as follows. Section 2 introduces briefly the concept of Banach space and the asymptotic correctness property. Section 3 introduces the EA probabilistic model and gives two asymptotic correctness theorems for EA processing chromosomes from an arbitrary Banach space. Section 4 illustrates few selected practical issues concerning the Gaussian mutation in R^n. Summary and conclusions are provided in Section 5.

2 Basic concepts

2.1 Banach space

A space S is called *Banach space* if it is linear, normed, metric, and complete. Linearity means that there are defined operators of addition $+ : S \times S \to S$ and multiplication $\cdot : R \times S \to S$. The following properties hold:

$$\forall \mathbf{x}, \mathbf{y} \in S \qquad \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \tag{1}$$

$$\forall \mathbf{x}, \mathbf{y} \in S, a \in R \qquad a \cdot (\mathbf{x} + \mathbf{y}) = a \cdot \mathbf{x} + a \cdot \mathbf{y} \tag{2}$$

$$\forall \mathbf{x} \in S \qquad \mathbf{x} = 1 \cdot \mathbf{x} \tag{3}$$

Elements of S are called vectors and denoted with boldface. There also exists an unique zero element $\mathbf{0} \in S$ such that

$$\forall \mathbf{x} \in S \qquad \mathbf{0} + \mathbf{x} = \mathbf{x} \tag{4}$$

We can define subtraction and negation as follows

$$\mathbf{d} = \mathbf{x} - \mathbf{y} \text{ iff } \mathbf{x} = \mathbf{y} + \mathbf{d} \tag{5}$$

$$-\mathbf{x} = \mathbf{0} - \mathbf{x} \tag{6}$$

Norm is a function $|| \cdot || : S \to R_+ \cup \{0\}$ with the following properties:

$$||\mathbf{0}|| = 0 \tag{7}$$

$$\forall a \in R, \mathbf{x} \in S \quad ||a \cdot \mathbf{x}|| = |a| \cdot ||x|| \tag{8}$$

$$\forall \mathbf{x}, \mathbf{y} \quad ||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \tag{9}$$

Metric is a function $\delta : S \times S \to R_+ \cup \{0\}$. In the Banach space the metric is defined with the use of the norm

$$\delta(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}|| \tag{10}$$

Banach space is complete, i.e. each convergent Cauchy series of points from S converges to a point from S.

The space is measurable if there exists a measure function $m : B(S) \to R_+ \cup \{0\}$, where $B(S)$ is a family of subsets of S, and

$$m(\emptyset) = 0 \tag{11}$$

$$\forall A, B \in B(S) \quad m(A \cup B) \leq m(A) + m(B) \tag{12}$$

$$m(A \cap B) \leq m(A) \text{ and } m(A \cap B) \leq m(B) \tag{13}$$

Let us additionally define a diagonal of a set as a function $d : B(S) \to R_+ \cup \{0\}$ such that

$$\forall A \in B(S) \quad d(A) = \sup_{\mathbf{x}, \mathbf{y} \in A} ||\mathbf{x} - \mathbf{y}|| \tag{14}$$

2.2 Asymptotic correctness

Consider a measurable Banach space S, an objective function $f : S \to R$ and a set of feasible solutions (feasible set) $D \subseteq S$. By $\mathbf{x}^*$ denote the point where f takes its global maximum f^* in D. An algorithm is asymptotically correct in D [5] if it generates the sequence of points $\mathbf{x}^t$ with the following property

$$\forall A \subseteq D, m(A) > 0 \qquad \lim_{T \to \infty} Prob \left\{ \exists t < T | \mathbf{x}^t \in A \right\} = 1. \tag{15}$$

Consider the sequence $\{\mathbf{y}^t\}$ defined as follows

$$\mathbf{y}^1 = \mathbf{x}^1 \tag{16}$$

$$\forall t > 1 \qquad \mathbf{y}^t = \begin{cases} \mathbf{x}^t & \text{if } f(\mathbf{x}^t) > f(\mathbf{y}^{t-1}) \\ \mathbf{y}^{t-1} & \text{otherwise} \end{cases} \tag{17}$$

Assume that $\mathbf{x}^*$ is not an isolated point, i.e.

$$\forall \varepsilon > 0 \qquad m(L_f(f^* - \varepsilon)) > 0, \tag{18}$$

where L_f is the level set defined as follows

$$L_f(M) = \{\mathbf{x} \in D, f(\mathbf{x}) > M\}. \tag{19}$$

The sequence $\{f(\mathbf{y}^t)\}$ will then converge to f^* with probability 1:

$$\forall \varepsilon > 0 \qquad \lim_{t \to \infty} Prob\left\{\mathbf{y}^t \in L_f(f^* - \varepsilon)\right\} = 1. \tag{20}$$

3 Asymptotic correctness of evolutionary algorithm

3.1 The algorithm

Consider a measurable Banach space S and the feasible set $D \subseteq S$, $m(D) < \infty$. Chromosomes are vectors from S, and the fitness function $f : S \to R$ should be maximized in D. The evolutionary algorithm maintains the base population $\mathbf{P}^t$ of chromosomes (t is the generation index). An initial base population $\mathbf{P}^0$ is created arbitrarily, although it is necessary that at least a single chromosome is contained by D.

The base population $\mathbf{P}^t$ is used to build a new offspring population $\mathbf{O}^t$. This is attained in two phases: reproduction (with or without crossover) and mutation of the reproduced individuals. After that, base population $\mathbf{P}^{t+1}$ for the new generation is selected from the populations $\mathbf{P}^t$ and $\mathbf{O}^t$. Perhaps the most popular choice is to set $\mathbf{P}^{t+1} = \mathbf{O}^t$.

It is assumed that there exists a method for generating a point from D. This method is used to fill in the whole population $\mathbf{P}^0$ in the EA initialization phase. It is also assumed that if $\mathbf{P}^t$ contains no feasible individuals, then at least one element from $\mathbf{P}^t$ is replaced by a point generated from D with the same procedure as used for the initialization.

3.2 Probabilistic model of EA

Operation of the EA can be described in probabilistic terms. The general idea is to treat the reproduction (with or without crossover) and mutation process as a random variable and to determine its distribution, called the *sampling distribution* (SD). The offspring population $\mathbf{O}^t$ contains chromosomes drawn independently from that random variable.

Reproduction. Let us number the individuals in $\mathbf{P}^t$ with integers $1,...,\mu$. The i-th element $\mathbf{x}_i$ of the base population $\mathbf{P}^t$ can be selected to reproduce with the probability $P_r(i|\mathbf{P}^t)$. The probability for selection is defined in the conditional way, since the contents of the whole base population influences its value. Note that some chromosomes from $\mathbf{P}^t$ may be identical, so is we want to describe the reproduction process in the search space S, we have to take this fact into account. Thus, the distribution of reproduction in the search space $P_r(\mathbf{x}|\mathbf{P}^t)$ satisfies the following condition

$$\forall A \subseteq S \quad Prob(\mathbf{x} \in A) = \sum_{\mathbf{x}_i \in \mathbf{P}^t, \mathbf{x}_i \in A} P_r(i|\mathbf{P}^t) \tag{21}$$

where $\mathbf{x}$ is the vector resulting from the reproduction.

Reproduced chromosomes undergo mutation either immediately, or after crossing them over. By p_c denote the probability that an individual from $\mathbf{O}^t$ is yielded by the reproduction+crossover+mutation cycle; obviously $1-p_c$ is the probability of mutating a chromosome immediately after the reproduction. Then the distribution of chromosomes

that undergo further reproduction is

$$P_r'(\mathbf{x}|\mathbf{P}^t) = p_c P_{rc}(\mathbf{x}|\mathbf{P}^t) + (1 - p_c)P_r(\mathbf{x}|\mathbf{P}^t) \tag{22}$$

where P_{rc} is the reproduction-with-crossover distribution.

Reproduction with crossover. Consider the $P_{rc}(\mathbf{x}|\mathbf{P}^t)$ distribution. Assume that two parental chromosomes $\mathbf{x}$ and $\mathbf{y}$ undergo crossing over and yield a single chromosome $\mathbf{c}$. If the role of crossover is to generate chromosomes sharing the parents' genetic material, then in a Banach space it is natural to assume that

$$||\mathbf{c} - \mathbf{x}|| + ||\mathbf{c} - \mathbf{y}|| = ||\mathbf{x} - \mathbf{y}||. \tag{23}$$

Note that above property holds for the "canonical" crossover schemes, e.g. arithmetical crossover in R^n or 1-point, k-point and uniform crossover in $\{0,1\}^n$.

We can use (22) and (23) to give a lower bound for the probability of reproduction with crossover. Observe that, since individuals are reproduced independently one from another, it is possible to select and reproduce twice the same individual which in turn results in $\mathbf{c} = \mathbf{x}$. If so, and if the crossed over parents are picked up on random according to the reproduction distribution $P_r(\mathbf{x}|\mathbf{P}^t)$, then

$$P_{rc}(\mathbf{x}|\mathbf{P}^t) \geq [P_r(\mathbf{x}|\mathbf{P}^t)]^2 \tag{24}$$

So, putting (24) into (22) we get the following bound

$$P_r'(\mathbf{x}|\mathbf{P}^t) \geq p_c[P_r(\mathbf{x}|\mathbf{P}^t)]^2 + (1 - p_c)P_r(\mathbf{x}|\mathbf{P}^t) \tag{25}$$

Note that $P_{rc}(\mathbf{x}|\mathbf{P}^t) > 0$ provided that $P_r(\mathbf{x}|\mathbf{P}^t) > 0$. Note also that there exist some points for which $P_{rc}(\mathbf{x}|\mathbf{P}^t) > 0$ although $P_r(\mathbf{x}|\mathbf{P}^t) = 0$. Surprisingly, for some points it may happen that $P_r'(\mathbf{x}|\mathbf{P}^t) > P_r(\mathbf{x}|\mathbf{P}^t)$.

Mutation. The mutation operation is defined independently from the base population contents. Mutation yields a new chromosome $\mathbf{y}$ by adding a random correction vector $\mathbf{m}$ to the chromosome $\mathbf{x}$ which undergoes mutation. Thus, if mutation is performed on the result of the reproduction $\mathbf{r}$ (with or without crossover), and $\mathbf{r}, \mathbf{m}$ are generated independently, the resulting chromosome

$$\mathbf{y} = \mathbf{r} + \mathbf{m} \tag{26}$$

is a random variable with the distribution $P_s(\mathbf{x}|\mathbf{P}^t)$ given by

$$P_s(\mathbf{x}|\mathbf{P}^t) = P_r'(\mathbf{x}|\mathbf{P}^t) * P_m(\mathbf{x}), \tag{27}$$

where '*' is the symbol of convolution. Note that the resulting distribution $P_s(\mathbf{x}|\mathbf{P}^t)$ is the sampling distribution.

3.3 Two theorems on EA asymptotic correctness

Let us now consider two theorems about the asymptotic correctness of an EA in a measurable Banach space. In the first theorem (after [8], generalized version of the theorem 4.2.2.3_1), mutation is assumed to have positive distribution over the whole feasible set.

Theorem 1 *Consider a measurable Banach space S and the feasible set $D \subseteq S$, $d(D) < \infty$. Assume that EA uses mutation (26) whose distribution satisfies the condition*

$$\exists r \geq d(D), \varepsilon > 0 \quad \forall A \subseteq K(r) \qquad Prob\left\{\mathbf{m} \in A\right\} \geq \varepsilon m(A) \tag{28}$$

where $K(r) = \{\mathbf{x} \in S, ||\mathbf{x}|| < r\}$. The evolutionary algorithm is asymptotically correct in D, regardless of the selection, crossover and the contents of the population $\mathbf{P}^0$.

The idea of proving is to observe the base population contains in each generation at least one feasible individual $\mathbf{x}^t \in D$ for which $P_r(\mathbf{x}^t|\mathbf{P}^t) \geq 1/\mu$. Since for each $C \subseteq D$ and for each $\mathbf{y} \in C$ we get $||\mathbf{x}^t - \mathbf{y}|| \leq d(D)$, then due to the definition of mutation (26) we get it holds

$$\exists \varepsilon > 0 \quad \forall C \subseteq D, \mathbf{x} \in D \qquad Prob\left\{(\mathbf{x} + \mathbf{m}) \in C\right\} \geq \varepsilon m(C) \tag{29}$$

and finally, taking into account the probability for reproducing $\mathbf{x}$, we conclude that

$$\exists \varepsilon > 0 \quad \forall C \subseteq D, t, \mathbf{P}^t \quad \exists \mathbf{x} \in \mathbf{P}^t \qquad Prob\left\{(\mathbf{x} + \mathbf{m}) \in C\right\} \geq \varepsilon m(C)/\mu \tag{30}$$

which proves the asymptotic correctness of the EA.

Note that Theorem 1 points to macromutations as to the essential mechanism allowing for the global properties of the EA, regardless of the selection type. This is somehow inconsistent with the evolution theory which explains the amazing adaptation properties as an effect of the cumulation of small mutations over generations rather than single macromutations (e.g. the discussion about the appearance of eyes)[3].

In addition there is a problem with Theorem 1 from the numerical point of view. In the R^n space, which is a well known Banach space, it is customary to use the zero mean Gaussian distribution to implement the mutation. Available random number generators only approximate the Gaussian distribution and it is practically impossible to generate a random value exceeding the range $[-5\sigma, 5\sigma]$. Thus one may come up to the situation when the assumptions of the theorem 1 are satisfied by the evolutionary algorithm and do not hold for its computer implementation. In order to overcome the above problems it is necessary to formulate the theorem which is less demanding when considering the mutation distribution.

Theorem 2 *Consider measurable Banach space S and the feasible set $D \subseteq S$, $d(D) < \infty$. Assume that EA is defined in the following way:*
  - *mutation operator is defined according to (26) and the condition is satisfied*

$$\exists r > 0, \varepsilon > 0 \quad \forall A \subseteq K(r) \qquad Prob\left\{\mathbf{m} \in A\right\} \geq \varepsilon m(A) \tag{31}$$

    where $K(r) = \{\mathbf{x} \in S, ||\mathbf{x}|| < r\}$,
  - *reproduction distribution (22) is positive defined*

$$\exists \varepsilon > 0 \quad \forall t \quad \forall i \in \{1, ..., \mu\}, \mathbf{x}_i \in D \qquad P_r(i|\mathbf{P}^t) > \varepsilon \tag{32}$$

  - *if crossover is used then it satisfies the property (23)*
  - *offspring population becomes the base population for the next generation $\mathbf{P}^{t+1} = \mathbf{O}^t$.*

The evolutionary algorithm is asymptotically correct in D, regardless of the contents of the population $\mathbf{P}^0$.

Note that Theorem 2 points to the cumulation of small changes (possibly leading temporarily to the fitness decrease) as to the mechanism crucial for the global properties of the EA.

The proof is based on the observation that the population $\mathbf{P}^t$ contains individuals resulting from k mutations of an individual from $\mathbf{P}^{t-k}$. Thus

$$\forall t \geq k \quad \exists \mathbf{y} \in \mathbf{P}^t, \mathbf{x} \in \mathbf{P}^{t-k} \qquad \mathbf{y} = \mathbf{x} + \sum_{i=1}^{k} \mathbf{m}_i \tag{33}$$

where $\mathbf{m}_i$ are k vectors driven independently from the random variable with the distribution P_m, and $k = \lceil d(D)/r \rceil + 1$ where r characterizes the distribution P_m in (31). It can be proved [1] that if a distribution $P(\mathbf{x})$ satisfies the property

$$\exists r > 0, \varepsilon_1 > 0 \quad \forall A \subseteq K(r) \qquad Prob\{\mathbf{x} \in A\} \geq \varepsilon m(A) \tag{34}$$

where $K(r) = \{\mathbf{x} \in S, \|\mathbf{x}\| < r\}$, then for the distribution

$$P_l(\mathbf{x}) = \underbrace{P_m * \cdots * P_m}_{k}(\mathbf{x}) \tag{35}$$

we get

$$\exists \varepsilon_2 > 0 \quad \forall A \subseteq K((k-1)r) \qquad Prob\{\mathbf{x} \in A\} \geq \varepsilon m(A) \tag{36}$$

Note that the position of $\mathbf{y}$ from (33) is a random variable with the distribution dependent on $\mathbf{x}$

$$P(\mathbf{y}|\mathbf{x}) = \underbrace{P_m * \cdots * P_m}_{k}(\mathbf{y} - \mathbf{x}) \tag{37}$$

If (31) holds and $d(D) < \infty$ then $k < \infty$ and, taking into account (34)–(36), we get

$$\exists \varepsilon > 0 \quad \forall A \subseteq K(d(D)), \mathbf{x} \in D \qquad Prob\{(\mathbf{y} - \mathbf{x}) \in A\} \geq \varepsilon m(A) \tag{38}$$

where $\mathbf{y}$ is given by (33). For each $\mathbf{x} \in \mathbf{P}^{t-k}$ the probability that there exists $\mathbf{y} \in \mathbf{P}^t$, such that (33) holds, is no less that a certain number $\varepsilon > 0$, due to (32) and (25). The latter observation, together with (38), proves the Theorem 2.

4 Practical issues

The above theorems give important time-asymptotic information about the dynamics of an EA. It would be interesting to compare what are the lower bound values for the probability of generating a point within a certain area of the feasible set D.

Consider an EA in R^1 and a feasible set D such that $d(D) < \infty$. Assume first that the mutation is performed according to a distribution uniform in the range $[-1, 1]$. This distribution does not satisfy the condition of the Theorem 1, though the asymptotic correctness can be proved using the Theorem 2.

Now assume that Gaussian mutation and tournament selection are used (tournament size equals 2). Assume also that infeasible individuals do not reproduce. According to

the assumptions of the Theorem 1, the sampling distribution can, in each generation, be bounded by

$$\varepsilon_1 = N_{0,\sigma}(d(D)) \tag{39}$$

where $N_{0,\sigma}(x)$ is the pdf of the normal distribution of mean 0 and standard deviation σ. After t generations, the probability that EA will generate at least one point in a set $A \subseteq D$ is no less than

$$p_1(t) = 1 - [1 - \varepsilon_1 m(A)]^t \tag{40}$$

If one follows the assumptions of the Theorem 2, and if the base population contains at least one feasible individual through all generations, the pdf of the sampling distribution after t generations can be bounded by

$$\varepsilon_2(t) = \mu^{-2t} N_{0,\sigma\sqrt{t}}(d(D)) \tag{41}$$

since the variance of t autoconvolutions is t times greater than the convolved distribution, and each feasible individual reproduces with the probability no less than μ^{-2}. On the other hand, if we start computations in generation $k < t$, we will get that in the same time a bound

$$\varepsilon_2(t) = \mu^{-2(t-k)} N_{0,\sigma\sqrt{t-k}}(d(D)) \tag{42}$$

Finally, the pdf of the sampling distribution is bounded by

$$\varepsilon_2(t) = \max_{k=0,\ldots,t-1} \mu^{-2(t-k)} N_{0,\sigma\sqrt{t-k}}(d(D)) \tag{43}$$

So, after t generations, the probability that EA will generate at least one point in a set $A \subseteq D$ is bounded by

$$p_2(t) = 1 - \prod_{k=1}^{t} [1 - \varepsilon_2(t) m(A)] \tag{44}$$

where $\varepsilon_2(t)$ is given by (43).

Let us investigate the bounds p_1 and p_2 computed according to both Theorems.

Assume $d(D) = 8$, $\sigma = 1$, $\mu = 5$, $m(A) = 1$. Plots of p_1 and p_2 versus the number of generations are provided in Fig.1. Extremely low values of both parameters can easily be observed. It can also be seen that after a few generation we have $p_2(t) \approx p_1(t)$.

One can expect that the population size μ should strongly influence the proportion $p_2(t)/p_1(t)$. This effect is illustrated in Fig.1. Plots have been generated assuming $d(D) = 8$, $\sigma = 1$ and $m(A) = 1$. It can be seen that for a small population size, the bound p_2 is significantly greater than p_1. This changes when $\mu > 10$. We can judge that the asymptotic correctness of EA with small populations can be explained mainly due to the accumulation of mutations, and with the large ones — due to macromutations.

Influence of the σ value on the $p_2(t)/p_1(t)$ seems obvious — the larger σ, the better bounds we get for macromutations and the less important is the accumulation of the mutations over successive generations. This effect is illustrated in Fig.2 assuming $d(D) = 8$, $\mu = 5$ and $m(A) = 1$. It was impossible to generate plots for $\sigma < 0.9$ due to extremely small $p_1(t)$ values. It seems that if σ is small in comparison with $d(D)$ (in our case $\sigma < 1.4$, i.e. $d(D) > 5.7\sigma$), macromutations are definitely not the mechanism that explains the asymptotic correctness.
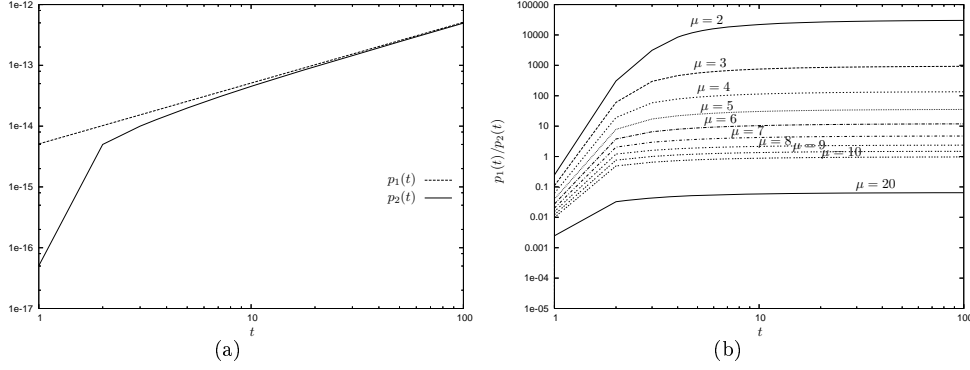
Figure 1. a) Values of $p_1(t)$ and $p_2(t)$ vs. t, b) Influence of the population size μ on the fraction $p_1(t)/p_2(t)$.
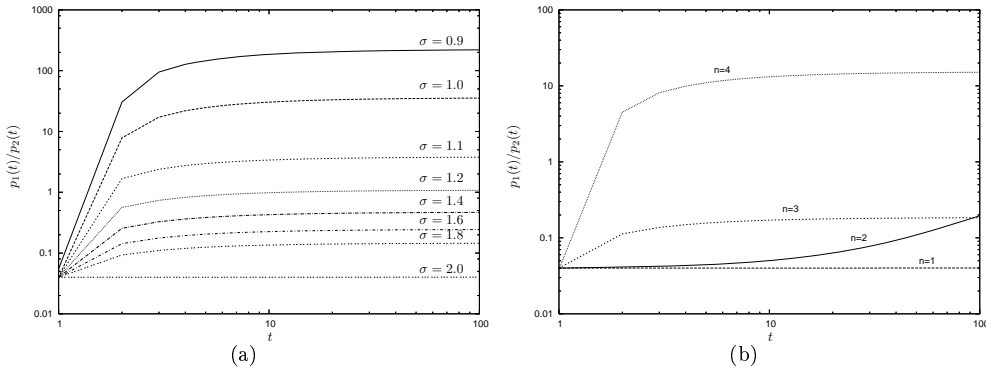

Figure 2. Influence on the fraction $p_1(t)/p_2(t)$ of the a) mutation range σ and b) search space dimensionality n.

According to the intuition we could expect that the role of macromutations will decrease with the dimensionality of the search space. To verify this let us compute $p_2(t)$ and $p_1(t)$ assuming that the evolution is performed in R^n, the feasible set is a hypercube with the edge of the length 2, so $d(D) = 2\sqrt{n}$, and $m(A) = 1$ (so A is a unit hypercube), and a population of size $\mu = 5$ is processed. Again it appears that with n increasing, the lower bounds obtained for macromutations vanishes faster than for the accumulation of small mutations (see Fig.2).

5 Summary and conclusions

The paper is devoted to certain aspects of the relationship between the abstract evolutionary algorithm and its computer implementation. Theorems for proving the asymptotic correctness of EA usually take an assumption that the mutation distribution is positive everywhere (macromutation based theorems). In this paper a new version of the asymptotic correctness theorem is presented with a weaker assumption about the mutation

distribution: it is assumed that there exists a ball of a certain radius in which the mutation distribution is positive (accumulation based theorem). It is however necessary that the selection is nonelitist and each individual, even the worst, may generate its offspring.

The accumulation based theorem and the macromutation based one use certain lower bounds for the probability of putting a point in an arbitrary subset of the feasible set. A simple illustrative example is used to investigate the relationship of the lower bound values for both theorems. It seems that the relationship depends on the conditions of an EA: if the mutation distribution is concentrated on a small-diagonal set, the population size is small, or the number of dimensions is high, the accumulation based theorem gives better values of the lower bound. Otherwise, the estimations due to the macromutation based theorem are better. One may conclude, that these observations imply the conditions, under which it is more advisable to rely either on large populations and macromutations, or on the accumulation of small mutations by a flexible small population.

Bibliography

[1] Jarosław Arabas. *Evolutionary computation as a search method in Banach space.* Oficyna Wydawnicza Politechniki Warszawskiej, 2005.

[2] Dipankar Dasgupta and Zbigniew Michalewicz. *Evolutionary Algorithms for Engineering applications.* Springer, 1997.

[3] Richard Dawkins. *Blind watchmaker.* Biblioteka Myśli Współczesnej. PIW, Warszawa, 1994.

[4] David Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* MIT Press, 1989.

[5] Reiner Horst and Panos Pardalos. *Handbook of global optimization.* Kluwer, 1995.

[6] Zbigniew Michalewicz. *Genetic Algorithms+Data Structures=Evolution Programs.* Springer, 1992.

[7] Günther Rudolph. Convergence of evolutionary algorithms in general search spaces. In *Proc. 3rd IEEE Conf. on Evolutionary Computation*, pages 50–54, 1996.

[8] Robert Schaefer. *Fundamentals of genetic global optimization.* Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, 2002.

[9] Robert Schaefer. Detailed evaluation of the schemata cardinality modification at the single evolution step. In *KAEiOG*, pages 143–148, 2004.

[10] Michael Vose. *The simple genetic algorithm.* MIT Press, 1999.