

Adaptive Reservoir Genetic Algorithm: Convergence Analysis

CRISTIAN MUNTEANU, AGOSTINHO ROSA
LaSEEB-ISR-Instituto Superior Técnico,
Universidade Técnica de Lisboa
Av. Rovisco Pais, 1-Torre Norte 6.21, 1049-001, Lisboa
PORTUGAL

Abstract: - This paper extends the theoretical analysis of the Adaptive Reservoir Genetic Algorithm (ARGA), a variant of a Genetic Algorithm (GA) proposed by the authors in [4]. We show that ARGA visits the global optimum after a finite number of iterations with probability one, regardless of the initialization of the population.

Key-Words: - Genetic Algorithms, convergence analysis, adaptive control of search.

1 Introduction

Genetic Algorithms (GAs) are stochastic search strategies that mimic the evolution of populations of individuals. GAs, in their classical outfit [3], [2] proved to be insufficient in coping with difficult optimization problems (like non-separable problems, highly epistatic, deceptive, or highly multi-modal). This is due mainly to the fact that GAs in the standard form lack the required adaptability (exploration/exploitation balance), being often prone to premature convergence [1]. In [4] we propose a novel model of a GA, that better adapts exploration/exploitation during time, based on a control mechanism where the control signal is the best fitness in the population, while the controlled variable is the population diversity through the mutation operator. The present paper gives a limit behavior in finite space and discrete time. Following a theorem from [6] it is shown that ARGA finds the global optimum with probability one, in finite time and regardless of the initialization of the population. Thus, ARGA is shown to be well behaved (finds the global optimum in finite time) both in theory and in practice as other papers by the authors have shown [4], [5].

2 ARGA: Limit Behavior In Finite Space And Discrete Time

2.1 ARGA structure

ARGA brings a novel mechanism for mutating the individuals in the population in conjunction with the

selection mechanism. In a standard GA each chromosome in the population can be mutated, depending on the probability of mutation [2]. In ARGA, we restrict mutations to a subpopulation of chromosomes, called *reservoir*. The reservoir has its individuals mapped onto a fixed population. The number of chromosomes in the mutant subpopulation (reservoir) is called *diameter* and is adapted during run. If there is no improvement in the best solution found during a certain number of generations, the diameter of the reservoir grows, in order to obtain a larger diversity in the population and to recast the search in a better niche of the search space. When this event occurs (i.e. an improvement beyond a certain threshold) the diameter of the reservoir is reset to the initial value. The algorithm is given in pseudo-code, as follows from Figure 1.

The adjustment of the reservoir's diameter $\Delta(t)$ is first done by comparing the best individual in the current generation t with the best individual in the previous generation $t-1$. If there is an improvement of the best fitness found beyond a certain threshold ϵ , the diameter is reset to its initial value Δ_0 . Otherwise, a constant rate $c > 0$ is added to $\Delta(t-1)$ and the integer part of the sum is taken to be the new reservoir's diameter $\Delta(t)$. If the reservoir becomes bigger than the size of the population, again the diameter is reset to its initial value Δ_0 .

ARGA ()

```
{
  -Start with an initial time counter  $t$ .
  -Initialise a random population  $P(t)$  within
  specific bounds.
  -Set the initial value of reservoir's diameter  $\Delta$ ,
  to  $\Delta_0$ .
  -Compute fitness for all individuals.
  while not done do
    -Increase the time counter.
    -Select parents to form the intermediate
    population by applying binary
    tournament.
    -Perform mutation on reservoir  $r(t)$ .
    {
      -Select reservoir  $r(t)$ , by choosing in a
      binary tournament the less fitted  $\Delta$ 
      individuals in the intermediate
      population.
      -Perform mutation on  $r(t)$  with a
      random rate between 0 and  $P_m$ .
      -Introduce mutants in the intermediate
      population.
    }
    -Perform (one-point) crossover on
    intermediate population with a rate  $P_c$ .
    -Form the population in the next
    generation by applying a  $k$ -elitist scheme
    to intermediate population.
    -Compute the new fitness for all
    individuals.
    -Adjust the diameter of the reservoir:
     $\Delta(t)$ .
  od
}
```

Figure 1. ARGA's structure.

Thus, ARGA acts like an adaptively controlled search mechanism where the control signal is given by the best fitness found in the current generation, and the end-controlled parameter is the diversity of the population, through the mutation operator. A typical dynamics of ARGA is given in Figure 2. It may be seen that as the best fitness doesn't increase for several generations in a row, the mean fitness of the population starts decreasing due to a larger exploration of the search space: the reservoir grows, more individuals are subject to mutations, there is a larger genotypic diversity in the population.

As a new better super individual is found, a takeover mechanism starts. The reservoir is reset to its initial value Δ_0 (usually a small integer value, such that $\Delta_0 \ll N$), less individuals are subject to mutations, and the population becomes less diverse (more homogenous), due to crossovers that together with selection and small scale mutations act like a convergence force to the already found best individual. ARGA, thus exploits the search space around the super individual. For example, in Figure 2, a super individual was found around generation 35 (having a fitness below 0.95), and the algorithm, cannot find a better one until generation 85 (an individual of fitness above 0.95). First, ARGA starts the exploitation around the individual with fitness below 0.95. As no better individual is found for several generations, at about generation 50, a mechanism comes into effect pushing ARGA from exploitation, more into exploration of the search space, until a new best individual is found.

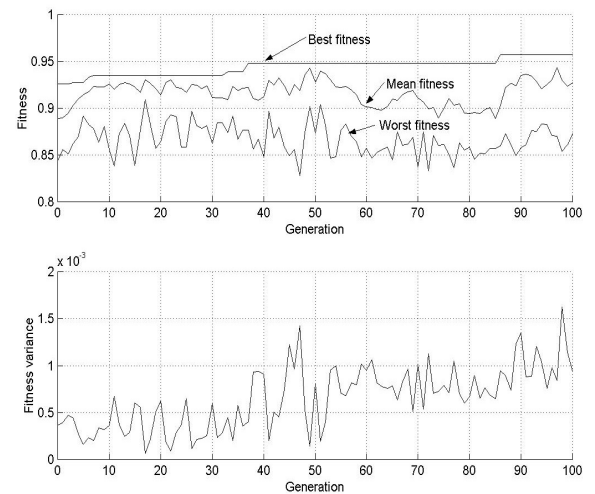


Figure 2. A typical dynamics of ARGA.

ARGA has the following specific parameters:

- k in the k -elitist scheme, is the number of the most fitted chromosomes that are automatically copied in the next generation.
- Δ_0 is the initial size or diameter of the reservoir r .
- ϵ is the threshold beyond which a variation of the best fitness is considered significant.
- c is a real value parameter that gives the rate at which $\Delta(t)$ grows when there is no significant improvement of the best fitness.

These are parameters specific to ARGA, but the actual number of parameters is bigger by adding the parameters in a standard GA: the size of the population N , the crossover rate P_c and the

maximum mutation rate: P_m . Details about ARGAs architecture, theoretical analysis of its structure and comparison with other similar strategies can be found in [4].

2.2 Convergence analysis

The convergence analysis for ARGAs uses Evolutionary Algorithms (EAs) convergence theorems given in [6]. Other convergence theorems applying Markov Chain analysis are summarized in [6] and [7]. We first review the main theoretical fact from [6] and then we apply the respective theory to ARGAs. Let $(x_1, x_2, \dots, x_N) \in c^N$ denote the population of N parents and c is the search set (the domain of the individuals). A general EA proceeds as follows: at first, R parents are selected to serve as mates for the recombination process: $\text{mat}: c^N \rightarrow c^R$ where $2 \leq R \leq N$. These individuals are then recombined by the procedure: $\text{reco}: c^R \rightarrow c$ yielding a partial offspring. Next, the mutation is applied to this offspring generating the complete offspring: $\text{mut}: c \rightarrow c$. After generating all M offspring in this manner, the selection procedures decides which of the offspring and possibly parents remain in the population to form the new generation of parents, in the next generation: $\text{sel}: c^Q \rightarrow c^N$ with $Q \geq N$. After formalizing these operations during a generation of the EA, several properties of the variation and selection operators are given in what follows:

Conditions:

- a) $\forall x \in (x_1, x_2, \dots, x_N): P\{x \in \text{reco}(\text{mat}(x_1, x_2, \dots, x_N))\} \geq d_r > 0$.
- b) For every pair $x, y \in c$ there exists a finite path $(x_1, x_2, \dots, x_q)$ of pairwise distinct points with $x_1 = x$ and $x_q = y$ such that: $P\{x_{i+1} = \text{mut}(x_i)\} \geq d_m > 0$ for all i between 1 and $q-1$.
- b') For every pair $x, y \in c$ holds:
 $P\{y = \text{mut}(x)\} \geq d_m > 0$.
- c) $\forall x \in (x_1, x_2, \dots, x_Q): P\{x \in \text{sel}(x_1, x_2, \dots, x_N)\} \geq d_s > 0$.
- d) Let $v_Q^*(x_1, x_2, \dots, x_Q) = \max \{f(x_i) : i = 1, \dots, Q\}$ denote the best fitness value within a population of Q individuals ($Q \geq N$). The selection verifies the condition:

$$P\{v_N^*(\text{sel}(x_1, x_2, \dots, x_Q)) = v_Q^*(x_1, x_2, \dots, x_Q)\} = 1.$$

The conditions above bear the following meaning: assumption (a) means that every parent may be selected for mating and is not altered by recombination with minimum probability $d_r > 0$,

where we might recognize that $d_r = 1 - P_c$ with P_c the probability to do crossover. Assumption (b) ensures that every individual can be changed to an arbitrary other individual by a finite number of successive mutations, while assumption (b') asserts the same but within a single mutation. Assumption (c) says that every individual competing for survival may survive with minimum probability $d_s > 0$. This assumption works for all selection strategies besides truncation selection. Assumption (d) makes sure that the best individual among the competitors in the selection process will survive with probability one. This assumption is valid for all elitist selection strategies.

Theorem:

If the assumptions (a), (b) and (c) are valid then the evolutionary algorithm visits the global optimum after a finite number of iterations with probability one, regardless of the initialization. If assumption (d) is valid additionally and the selection method chooses from the parents as well as from the offspring then the evolutionary algorithm converges completely and in mean to the global optimum regardless of the initialization. For a demonstration of the theorem see [6].

Let us apply the theorem to ARGAs by showing that it obeys the assumptions in the first part of the theorem. Let us define the sequence $(X_t) : t \in \mathbb{N}$ to be a sequence of populations generated by some evolutionary algorithm and let:

$F_t = \max \{f(x_{t,1}), \dots, f(x_{t,N})\}$ denote the best objective value of the population at generation t .

Let $f^* = \max \{f(x) : x \in c\}$ with $f : c \rightarrow \mathbb{R}$, be the global optimum (maximum) for the function f . We define $T = \min \{t \geq 0 : F_t = f^*\}$ as the first hitting time of the global optimum. The first part of the theorem actually shows that $P\{T < \infty\} = 1$ regardless of the initialization. As ARGAs applies crossover with some probability $0 < P_c < 1$ we have that:

$$\forall x \in (x_1, x_2, \dots, x_N): P\{x \in \text{reco}(\text{mat}(x_1, x_2, \dots, x_N))\} \geq d_r = 1 - P_c > 0.$$

Therefore, ARGAs satisfies condition (a). In practice, ARGAs has been used before [4] with $P_c = 1$, however the convergence is guaranteed theoretically only with crossover probabilities lower than 1 (with $1 - P_c$ made arbitrary small). We next check assumption (c). As ARGAs uses a combination of tournament selection and k -elitism, each member of the population has a non-zero probability to be selected, that is:

$\forall x \in (x_1, x_2, \dots, x_Q): P\{x \in \text{sel}(x_1, x_2, \dots, x_N)\} \geq d_s > 0$.

To check assumption (b) we have two possible hypothesis:

bA) all elements x_i ($i=1$ to q) on the path that connects through successive mutations, any two individuals $x, y \in c$, lie within the reservoir. That means that all elements in the path are also in the reservoir. Therefore, they are subjected to mutation done with probability $p_m > 0$. Equivalently, we have that:

$P\{x_{i+1} = \text{mut}(x_i)\} = p_m > 0, \forall i = \overline{1, \dots, q}$. Therefore, in this case condition (b) is fulfilled.

bB) if $\exists x_i \notin r$ (i.e. there is at least one element in the path that connects x to y , that does not belong to the reservoir). In this case the reservoir grows such that after $T_\lambda < \infty$ generations we have $x_i \in r(T_\lambda)$. Therefore, we must wait for T_λ generations until the reservoir grows such as to contain x_i . Until this moment x_i remains unaltered with the probability $(P_s(x_i))^{T_\lambda} (1 - P_c(x_i))^{T_\lambda} > 0$ where we used the fact that ARGGA fulfills: $P_s(x_i) > 0$ and $P_c(x_i) < 1$, as discussed before. We get that:

$$P\{x_{i+1} = \text{mut}(x_i)\} = p_m \cdot P_s^{T_\lambda}(x_i) \cdot (1 - P_c(x_i))^{T_\lambda+1} > 0$$

where we used that $0 < p_m < P_m$, P_m being a positive parameter of ARGGA. To consider that ARGGA fulfills condition (b) we still have to show that $T_\lambda < \infty$. Suppose $T_\lambda = \infty$. This means the reservoir never grows. The reservoir not growing means that in each generation a better fitness is discovered. It follows that the sequence $\{F_t\}_{t \geq 0}$ is *strictly increasing* with time. However, as ARGGA works on finite space, it means that the set $\{f(x): x \in c\}$ is also *finite*. Thus, the reservoir cannot grow for every generation *ad infinitum*. We thus have that $T_\lambda < \infty$. Showing that ARGGA fulfills conditions (a), (b), (c) and applying the theorem given before, we deduct that *ARGGA visits the global optimum after a finite number of generations with probability one, regardless of the initialization*.

Two minor modifications have to be made to the original version of ARGGA in [4], in order that the convergence theorem be applied. That is: $P_c < 1$ and $p_m > 0$, instead of $0 \leq P_c \leq 1$ and $0 \leq p_m \leq 0.5 = P_m$, like in the initial version [4].

3 Conclusions

ARGGA has shown good search behavior both on test

problems [4] and on real-world applications like the Brain Computer Interface (BCI) [5]. In this paper, we further investigate the efficient behavior of ARGGA by showing that it finds the global optimum, with probability one, regardless of initialization. However, this property is a macroscopic property being too general to give insight of the way convergence is achieved: convergence dynamics, parameter influence on dynamics. A microscopic approach is needed to shed light on these matters. For future work we will concentrate our attention towards a microscopic analysis, however we are aware of the fact that this analysis might be problem-dependent, and difficult for problems for which the search landscape is not known beforehand.

References

- [1] Bäck T., Fogel D. B., and Michalewicz Z. eds. *Handbook of Evolutionary Computation*. IOP-Oxford University Press., 1997.
- [2] Goldberg D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1975.
- [3] Holland J. *Adaption in Natural and Artificial Systems*. Ann Arbor, 1975.
- [4] Munteanu C., Lazarescu V. Global Search Using a New Evolutionary Framework: The Adaptive Reservoir Genetic Algorithm, *Complexity International*. Life Science Publications/IOS Press, vol. 5, 1998.
- [5] Obermaier B, Munteanu C., Rosa A., Pfurtscheller G. Genetic Algorithm improved Hidden Markov Model Selection in offline Brain-Computer Interface Classification Task, accepted for publication in IEEE Trans. on Systems, Man and Cybernetics: Part C.
- [6] Rudolph G. Finite Markov Chain Results in Evolutionary Computation: A Tour d'Horizon. *Fundamenta Informaticae*, IOS Press, 1998, 1-22.
- [7] Vose M. *The Simple Genetic Algorithm*. MIT Press, 2000.