

Genetic Algorithms with Fitness- & Diversity-guided Adaptive Operating Probabilities and Analyses of its Convergence

Li Meiyi Cai Zixing Sun Guoyun

(College of Information Science & Engineering, Central South University, Changsha, China 410083)

meiy_li@yahoo.com.cn

Abstract The paper has analyzed global convergence properties of adaptive genetic algorithms combining adaptive probabilities of crossover and mutation with diversity-guided crossover and mutation. By means of homogeneous finite Markov chains, it is proved that AGAD, which is present in this paper, and GAD (genetic algorithms with diversity-guided mutation) maintaining the best solution converge to the global optimum, which is the main contributions of this paper. The performance of AGA (adaptive genetic algorithms with adaptive probabilities of crossover and mutation), GAD and AGAD in optimizing several unimodal and multimodal functions has been compared. For multimodal functions, the AGAD converges to the global optimum for fewer generations than AGA and GAD, and it hardly has premature convergence.

Key Diversity-guided mutation; Adaptive genetic algorithm; Markov chain analysis; Global convergence.

1. Introduction

It is known that the performance of the genetic algorithms (GAs) is dependent upon the operator probabilities used. Adaptation of operator probabilities makes the genetic algorithm a more effective optimizer. By adapting operator probabilities we can get benefit: (1) Increasing the quality of solutions obtained. (2) Allowing the GAs to find a solution of a given quality more quickly. Therefore, by employing some methods, some researchers attempt to automatically adjust the operator probabilities as the genetic algorithm runs according to the quality of solutions^[1,2,3], in which the quality of a solution determines its operator probabilities. It is also known that the premature convergence is a major problem in GAs and adaptive genetic algorithms may lead to premature convergence^[4], in order to overcome it, the term ‘diversity’ is employed. Diversity is undoubtedly closely related to the performance of

GAs, especially when attempts are made to avoid premature convergence and escaping local optima. Only a few papers^[5,6,7] have used diversity measure to control the search direction of evolutionary algorithms, and global convergence properties of genetic algorithms with diversity-guided crossover and mutation have hardly been discussed.

We make an attempt to mix adaptive crossover and mutation of Srinivas^[1] and diversity-guided crossover and mutation in order to get the balance between avoiding premature convergence and converging more quickly. We created AGAD, and proved that AGAD and GAD maintaining the best solution converges to the global optimum by homogeneous finite Markov chains in the paper, when diversity-guided mutation probabilities are bigger than 0, but AGA not always do so. In section 5 the performance of AGA, GAD and AGAD has been compared in optimizing several unimodal/multimodal functions. For multimodal functions, AGAD performs significantly better than AGA and GAD.

2. Preliminary

We make the assumption that genetic algorithms are used to tackle static optimization problem in which there are N individuals within the population noted as binary string of fixed length l :

$$g_{i1}g_{i2}\wedge g_{il} \quad (g_{ij}=0,1), i=1,2,\wedge N, \quad (1)$$

and correspond fitness values $\{f_i | 0 \leq f_i < \infty, i=1,2,\wedge N\}$.

Definition 1 Let Z_t be a sequence of random variables representing the best fitness within a population represented by state i at step t , and f^* is the global optimum of problem. if

$$\lim_{t \rightarrow \infty} p(Z_t = f^*) = 1, \quad (2)$$

then, it is called that a genetic algorithm converges to the global optimum.

In implement of GAs the convergence critical or condition of halting execution often is defined to be N_{gen} generations without best fitness improvement.

The convergence speed is the generations or halt generations that GAs has run before the convergence critical hold. In this paper, let $N_{gen} = 400$ for 15-dimension and $N_{gen} = 80$ for 2-dimension in function optimal problems.

Definition 2 A square matrix $A = (a_{i,j})_{n \times n}$ is said (1) nonnegative, if $a_{i,j} \geq 0$, $i, j \in \{1, \Lambda, n\}$; (2) positive, if $a_{i,j} > 0$, $i, j \in \{1, \Lambda, n\}$;

A nonnegative matrix A is said to be (3) primitive, if $\exists k > 0$, such that A^k is positive; (4)

stochastic, if $\sum_{j=1}^n a_{i,j} = 1$ $i \in \{1, \Lambda, n\}$;

A stochastic matrix A is said to be column allowable, if it has at least one positive entry in each column.

Lemma 1^[9] Let C, M and S are stochastic matrices, where M is positive and S is column allowable. Then the product CMS is positive.

Theorem 1 Let C_a, C, M, M_a and S are stochastic matrices. If M is positive, S is column allowable and M_a diagonally positive. Then the product $P = C_a C M M_a S$ is positive.

Proof Let $M = (m_{i,j})_{n \times n}$, $M_a = (p_{i,j})_{n \times n}$ and $M_0 = M M_a = (u_{i,j})_{n \times n}$. Since (1), we have

$$u_{i,j} = \sum_{k=1}^n m_{i,k} p_{k,j} \geq m_{i,j} p_{j,j} > 0.$$

So, M_0 is positive.

Note that the product of stochastic matrices is again a stochastic matrix, According to Lemma 1,

$P = C_a C M M_a S = C M_0 S$ is positive. Hence, the proof is completed.

3. Modified adaptive genetic algorithm AGAD

The modified genetic algorithm with adaptive crossover, mutation probabilities and diversity-guided mutation (AGAD) can be sketched as Fig.1. Adaptive crossover and mutation probabilities of individuals (denoted as i and j) within the population are $p_c(i, j)$ and $p_m(i)$

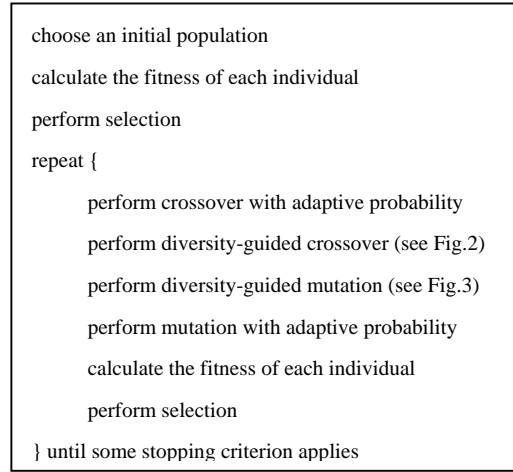


Fig.1 AGAD

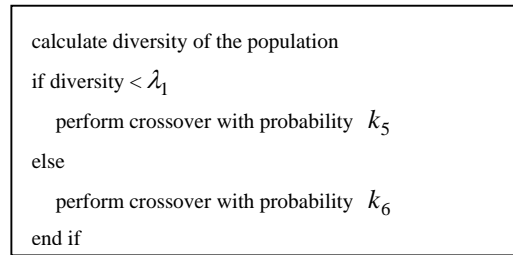


Fig.2 perform diversity-guided crossover

respectively, which are determined by individual fitness evaluations within the current population. The famous calculating method of $p_c(i, j)$ and $p_m(i)$ which is introduced by Srinvas^[1] is as follows

$$p_c(i, j) = \begin{cases} k_1 \cdot \frac{f_{\max} - f'}{f_{\max} - f_{avg}}, & f' \geq f_{avg} \\ k_2, & f' < f_{avg} \end{cases} \quad (3)$$

$$p_m(i) = \begin{cases} k_3 \cdot \frac{f_{\max} - f}{f_{\max} - f_{avg}}, & f \geq f_{avg} \\ k_4, & f < f_{avg} \end{cases} \quad (4)$$

Where $0 < k_1, k_2 \leq 1, 0 < k_3, k_4 < 1$, f_{avg} and f_{\max} are respectively the average fitness value and the maximum fitness value of the population, and f' is the larger of fitness values of the individuals i and j to be crossed. f is the fitness value of the individual i to be mutated.

Other papers^[2, 3] have developed Srinvas's adaptive genetic algorithm by means of modifying the formulae

```

calculate diversity of the population
if diversity <  $\lambda_2$ 
    perform mutation with probability  $k_7$ 
else if diversity <  $\lambda_3$ 
    perform mutation with probability  $k_8$ 
else
    perform mutation with probability  $k_9$ 
end if

```

Fig.3 perform diversity-guided mutation

(3) and (4).

Their adaptive genetic algorithms can increase the converge speed, but at least there are two shortage: (1) Crossover probability of the best individuals is very small (close to zero) so, the building-blocks of the best/better individuals do not spread out. (2) They often lead to premature convergence^[4].

In order to overcome premature convergence we use a diversity measure to alternate between exploring and exploiting behavior. A diversity measure for N-dimensional numerical problems is defined as:

$$diversity = 1 + \frac{1}{l} \sum_{j=1}^l p(j) \ln p(j), \quad (5)$$

where $p(j) = \frac{1}{|S| \cdot N} \left(\sum_{i=1}^N (g_{ij} - \bar{g}_j)^2 \right)^{1/2}$, $|S|$ is the

length of the diagonal in the search space $S \subset R^l$, and \bar{g}_j is the j 'th value of the average point, N and l are the size of population and the length of a individual respectively.

Genetic algorithms^[8] put stress on the crossover as the primary operator to search the space. Under crossover operators genetic algorithms exploit the area determined by the population in the search space. When probabilities of crossover operators are high, exploitations are more efficient.

The pseudo-code for "perform diversity-guided crossover" in the AGAD is listed in Fig. 2, and the pseudo-code for "perform diversity-guided mutation" in the AGAD is listed in Fig. 3, where $0 < \lambda_1 < 1$, $0 < \lambda_2 < \lambda_3 < 1$, $0 < k_5 < k_6 \leq 1$, $0 < k_9 < k_8 < k_7 < 1$ and $k_9 > 0$ is almost 0.

4. Markov Chain Analysis of AGAD and GAD

The AGAD or GAD can be described as a Markov chain over a finite state space $\Lambda = \{0,1\}^{l \cdot N}$ of cardinality

$|\Lambda| = l \cdot N$. The probabilistic changes of the individuals within the population caused by AGAD or GAD are captured by the transition matrix P , which can be decomposed in a natural way into a product of stochastic matrices $P = C_a C M M_a S$ for AGAD and $P = C M S$ for GAD, where C_a, C, M, M_a and S describe the intermediate transitions caused by crossover with adaptive probabilities, crossover with diversity-guided crossover probabilities, mutation with diversity-guided mutation probabilities, mutation with adaptive probabilities and selection, respectively. This leads to:

Theorem 2 The transition matrix of the AGAD with

diversity-guided mutation probability $0 < p_m < 1$,
 adaptive mutation probability $0 \leq p_m(i) < 1$ and
 adaptive crossover probability $p_c(i, j) \in [0, 1]$ and
 proportional selection is primitive.

Proof The crossover operators may be regarded as random total functions whose domains and ranges are Λ , i.e. each state of Λ is mapped probabilistically to another state. Therefore, C_a and C are stochastic. The same holds for the other operators and their transition matrices. Because the mutation operator is applied independently to each gene-bit in the population the probability that state i becomes state j after diversity-guided mutation can be aggregated to

$$m_{ij} = p_m^{H(i,j)} (1 - p_m)^{l-H(i,j)} > 0,$$

where H_{ij} denotes the Hamming distance between the binary representations of state i and state j .

After mutation with adaptive probability, the diagonal elements in M_a are

$$m_{ii}^a = (1 - p_m(i))^l > 0,$$

Thus M is positive, and M_a diagonally positive.

Since S is column allowable (the details see [10]), by Lemma 1, $P = C_a C M M_a S$ is positive. Since every positive matrix is primitive, the proof is completed.

Corollary 1 The AGAD is an ergodic Markov chain, i.e., there exists a unique limit distribution for the states of the chain with nonzero probability to be in any state at any time regardless of the initial distribution

If M is replaced with MM_a in the theorem 6 and 7 of the paper [10], the proof of the theorem 6 and 7 of the paper [10] hold. So we have following result.

Theorem 3 The AGAD maintaining the best solution found over time after/before selection converges to the global optimum.

If all the adaptive probabilities $k_1 = 0, k_2 = 0, k_3 = 0, k_4 = 0, k_5 = k_6$ and $k_9 > 0$, AGAD becomes into GAD. From Theorem 3 we have

Corollary 2 The GAD maintaining the best solution

found over time after/before selection converges to the global optimum.

As shown in [11], although AGA is a Markov chain over a finite state space $\Lambda = \{0, 1\}^{l \cdot N}$, when not maintaining the best solution found over time after/before selection, adaptive genetic algorithms of Srinivas^[1] don't converge to the global optimum. But when maintaining the best solution found over time after/before selection, since the transition matrix of adaptive genetic algorithms of Srinivas^[1] isn't positive (the mutation probability of the best individual is 0), we can't get result of convergence to the global optimum from the finite Markov chain theorem.

5. Experiments and Analyses

In this section, we discuss the experiments that we have conducted to compare the performance of the AGAD. For this purpose we will employ one unimodal and two nonunimodal functions with varying complexities as follows:

Quadric: A unimodal function with significant interaction between its variables. The global maximum is located at $X = 0$, so that $f(X) = 0$.

$$f(X) = -\sum_{i=1}^n \left(\sum_{j=1}^i x_j^2 \right)$$

Ackley: A multi-modal function with deep local maximum, which is $X = 0$, with $f(X) = 0$.

$$f(X) =$$

$$20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) + \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) - 20 - e$$

Rastrigin: A multi-modal version of the Spherical function, characterised by deep local maxima arranged as sinusoidal bumps. The global maximum is $f(X) = 0$, where $X = 0$. The variables of this function are independent.

$$f(X) = \sum_{i=1}^n (-x_i^2 + 10 \cos(2\pi x_i)) - 10$$

Table I lists the parameter settings for the functions in the benchmark suite.

Table 1 FUNCTION PARAMETERS

Function	Dim	Domain	Threshold
Ackley	15	$-30 \leq x_i \leq 30$	3.0
Rastrigin	15	$-5.12 \leq x_i \leq 5.12$	1.5
Quadric	15	$-100 \leq x_i \leq 100$	0.02

We use the genetic algorithm with simple crossover, binary mutation and proportional selection. The parameters of algorithms in this paper are set as follows. For the functions f_1, f_2 and f_3 , we used population of 100 and $k_1 = k_2 = 0.8$, $\lambda_1 = 0.3$, $\lambda_2 = 0.01$, $\lambda_3 = 0.25$, $k_5 = 0.3$, $k_6 = 0.85$, $k_7 = 0.8$, $k_8 = 0.03$, and $k_9 = 0.00001$, $k_3 = k_4 = 0.5$ for AGA and $k_3 = k_4 = 0.3$ for AGAD, since AGAD has yet diversity-guided mutation.

Let us discuss the genetic algorithm with diversity-guided mutation (GAD), in which the crossover probability is fixedly $p_c = 0.8$, and the mutation

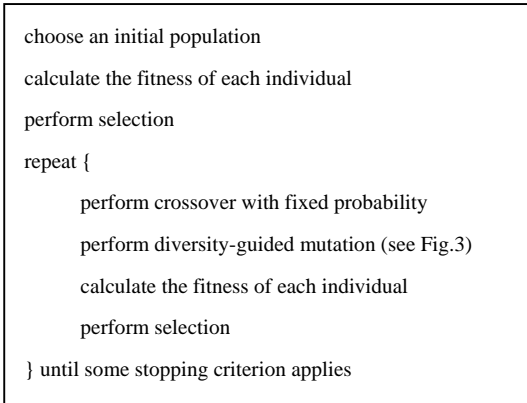


Fig.4 GA with diversity-guided mutation

probability is determined as in Fig.4. It has been mentioned that this algorithm can overcome premature convergence, but there are some shortages. Since lower the diversity is, more the exploitation for local area is, faster convergence to minima is, but more easily premature convergence happens. Otherwise, higher the

diversity is, more the exploration to global area is, and slower convergence to minima is. Of course the genetic algorithm with diversity-guided mutation can efficiently maintain the diversity of population.

If test function is non-unimodal, GA with only fixed mutation probability often suffers from premature convergence, and high diversity can help GA to escape from local optimum, exploring global optimum. AGA also suffers from premature convergence, but convergence is faster than GA with only fixed mutation probability. So, we present AGAD combining AGA and GAD.

Let us make some performance comparison of three algorithms. Table II presents average halt generations obtained by running the various algorithms until they discovered an objective function value below the corresponding value listed in the 'threshold' column of Table I. All experiments consisted of 50 runs. These results give an indication of the robustness as well as the rate of convergence of the three algorithms on the various benchmark functions.

Table 2 COMPARE PERFORMANCE

Function \ Algorithm	AGA	GAD	AGAD
Ackley	3753.51	31966.89	2655.26
Rastrigin	4711.73	3961.75	3348.96
Quadric	3342.03	3617.92	3482.40

The results listed in Table II clearly show that the AGAD outperforms the other algorithms, since the AGAD converges to the global optimum for fewer generations than AGA and GAD. Better fitness will be obtained as shown in Table 3 if the AGAD run until it stagnated, which is here defined to be 400 generations without fitness improvement. Note that all functions were tested using 15-dimensional search spaces. All experiments for every function consist of 20 runs respectively and the results of average best fitness are list in Table III on the three benchmark problems. Thanks to the diversity, AGAD and GAD are able to escape local optimum, and thanks to adaptive crossover

and mutation AGAD can converge more quickly. The last function, Quatric, which is modified Spherical function, is unimodal, but it has significant interaction between variables. So it is not easy to solve it. Conclusively, AGAD performs significantly better than AGA and GAD.

Table 3 AVERAGE BEST FITNESS

Algorithm Function	AGA	GAD	AGAD
Ackley	$5.165E^{-1}$	$1.160E^{-2}$	$2.437E^{-2}$
Rastrigin	3.336	$4.635E^{-2}$	$1.951E^{-2}$
Quadratic	$1.228E^{-3}$	$7.324E^{-3}$	$4.063E^{-3}$

6. Results

This paper analyzed the global convergence properties of AGA, GAD and AGAD, and it has been proved that AGAD maintaining the best solution found over time converge to the global optimum by means of homogeneous finite Markov chains.

AGAD combines the advantage of the two algorithms AGA and GAD, in which AGA used individual's level adaptive mutation probability, and GAD used population's level adaptive mutation probability. AGA converges fast, but often get into local minima; GAD can efficiently overcome premature convergence, but the convergence speed is limited under the thresholds of the diversity.

Some experiments in optimizing several unimodal and multimodal functions showed that AGAD has significantly faster convergence than AGA and GAD.

Combining diversity-guided mutation with other genetic operators converging faster, we will get better balance between overcoming premature convergence and convergence speed, and improve performance of GA, which is a meaningful investigation to do in the future.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 60234030).

References

- [1] Srinivas M and Patnai L M. Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms. *IEEE Trans.Syst.,Man,and Cybern.*,1994,24(4):656-666
- [2] ZONG Jing-qun. A Kind of Mixture Adaptive Genetic Algorithm and Analysis of Property. *Systems Engineering-Theory & Practice*, 2001 21(4): 14-18 (in Chinese)
- [3] Zhou Ming, Sun Shu Dong. *Genetic Algorithms: Theory and Applications*. Beijing: National Defence Industry Press, 1999.6 (in Chinese)
- [4] Rudolph G. Self-Adaptive Mutations May Lead to Premature Convergence. *IEEE Transactions on Evolutionary Computation* 1998, 5(4): 410-414
- [5] Burke E, Gustafson S, Kendall G. A Survey and Analysis of Diversity Measures in Genetic Programming. Eds: Morgan Kaufmann, *Proceedings of the Genetic and Evolutionary Computation Conference 3002 (GECCO 3002)*, New York, USA, 9-13 July 3001, 716-723.
- [6] Burke E, Gustafson S, Kendall G and Krasnogor N. Advanced Population Diversity Measures in Genetic Programming. In *Proceedings of Parallel Problem Solving from Nature Conference, 7th International Conference*. Granada, Spain. Sept. 3002. LNCS 2439, 341-350
- [7] Shimodaira H. A Diversity Control Oriented Genetic Algorithm (DCGA): Development and Experimental Results. In: *Proceedings of the Genetic and Evolutionary Computation Conference*, Eds: Banzhaf W, Daida J, Eiben A E, Garzon M H, Honavar V, Jakiela M and Smith R E, 1999, (1): 603-611.
- [8] Goldberg D E. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading MA, 1989.
- [9] Iosifescu M. *Finite Markov Processes and Their Applications*. Chichester: Wiley 1980: 123-126.
- [10] Rudolph G. Convergence Properties of Canonical Genetic Algorithms. *IEEE Transactions on Neural Networks*, 1994, 5(1): 96-101.
- [11] He Lin, Wang Ke Jun, Li Guo Bin and Jin Hong Zhang. The Discussion about the Paper "the Analysis of Global Convergence and Computational Efficiency for Genetic Algorithm". 2001, 18(01):141-144 (in Chinese)